Contents

Research Articles

Organon F 32 (2) 2025: 116–149 ISSN 2585-7150 (online)
https://doi.org/10.31577/orgf.2025.32201

RESEARCH ARTICLE

Internal Object Relations and the Logic of the Unconscious

Andriy Vasylchenko*

Received: 5 December 2023 / Revised: 27 November 2024 / Accepted: 19 March 2025

Abstract: This article explores internal objects as a kind of non-existent and analyses the role of internal objects and internal object relations in unconscious thinking. My aim is to outline a project of 'logic of the unconscious' based on object relations theory. Drawing on the approaches of Melanie Klein, Susan Isaacs, Joseph Sandler, Thomas Ogden, and Linda Brakel, I clarify unconscious thinking in terms of the notions of primary process and phantasy and explore the nature of internal objects as background phantasy figures. In addition, using psychoanalytically relevant clinical examples, I analyse condensation, displacement, and transference as primary mental operations and suggest that considering internal objects as abstract particulars helps to clarify their role in unconscious thinking. Finally, I formulate a project of the logic of the unconscious as a logic of object relations.

Keywords: Internal objects; logic of the unconscious; non-existent objects; object relations theory; phantasy; primary process.

* Czech Academy of Sciences
https://orcid.org/0000-0002-8531-656X
Department of Logic, Institute of Philosophy of the Czech Academy of Sciences, Jilská 1, Prague 1, 110 00, Czech Republic
an_vasylchenko@ukma.edu.ua

1. Introduction

The problem of the ontology of non-existent objects is a long-standing philosophical issue that has been discussed since Parmenides and Plato. In modern times, it was reformulated by Franz Brentano as the problem of "intentional in-existence" (Brentano 1995, 68) and re-emerged in the twentieth-century analytic philosophy after Alexius Meinong introduced his theory of objects (Meinong 1904) and Bertrand Russell published his critique of the theory (Russell 1905). The problem was addressed by W. V. O. Quine in "What There Is" (Quine 1948) and Rudolf Carnap in "Meaning and Necessity" (Carnap 1956). A review and comparative analysis of different approaches to the problem of non-existents, including most recent developments in analytic metaphysics, can be found in (Berto and Plebani 2015) and in a special issue of "Monist" on metaontology (Berto et al. 2014). Paradigmatic examples of non-existent objects include fictional objects (e.g. Sherlock Holmes or the Holy Grail), impossible objects such as the round square, and purely intentional objects (see Bebee et al. 2011, 148). More generally, the domain of non-existent objects encompasses all objects that do not exist spatiotemporally, including various kinds of abstracta.

The problem of the intentional in-existence of merely intentional objects has to do with the fact that "sometimes people imagine, desire, or fear things that do not exist" (Reicher 2022, 2). For example, "if my nephew is scared of the monster under his bed, the monster is the object of his fear but it does not exist" (Bebee et al. 2011, 148). Merely intentional objects have been studied extensively in the psychological sciences under the name of mental objects (see, e.g., Perlow 1995). An important type of mental object is what are called internal objects: internalised and transformed figures of our previous meaningful relationships. Most psychoanalysts and psychotherapists today accept some form of object relations theory, according to which the person's inner world is structured by an interaction between the person and her internal objects. However, despite their crucial role in psychotherapeutic theory and practice and their enormous influence on the way we perceive and act (see Auchincloss and Samberg eds. 2012, 175–178; Svrakic and Zorumski 2021, 4), internal objects have not yet been so far subjected to philosophical scrutiny as 'non-existents'.

What makes internal objects an elusive subject of study is that, on the one hand, they are formed on the basis of subjective experience, and, on the other, as unconscious formations, they often remain beyond the reach of first-person conscious mind. Fortunately, we can understand the dynamics of the formation and development of internal objects by analysing their role in unconscious (i.e., primary process-based) intentionality. Understanding mental objects means grasping how they function in the mental operations in which they are involved. Operations involving internal objects have been studied within theoretical psychoanalysis; the corresponding field of research, which explores the peculiarities of 'unconscious thinking', is known as 'the logic of the unconscious'. If we engage with these studies, we will find that primary process-based mental operations, such as transference, displacement, condensation, and the like, involve internal objects that can have different kinds of in-existence: being perceived, imagined, feared, desired, and so on (see section 8 below). Many of these operations can be modelled using contemporary non-classical logic (see Vasylchenko 2015), which allows us to elaborate on a logical theory of internal objects, to clarify their nature, and ultimately to include them in debates about what kind of being, if any, non-existents can have.

In view of their psychological significance, the logical challenges they pose, and the metaphysical possibilities they might open up, internal objects are reasonable objects of philosophical inquiry, perhaps no less diverse and intriguing than, say, the fictional objects or unobservable entities of the theoretical sciences. The integration of internal objects into philosophical logic and analytic philosophy has been prepared by recent developments in the analytic theory of intentionality, notably Graham Priest's logic of intentionality (Priest 2005) and Linda Brakel's theory of primary psychological attitudes (Brakel 2009). The aim of this article is to analyse internal objects as a kind of non-existent, to clarify their role in unconscious thinking, and to elaborate a project of a formal 'logic of the unconscious' based on object relations theory. I consider the realisation of this aim to be a first step towards the reception of internal objects in analytic metaphysics.

2. Logic of the Unconscious

Contemporary developments in the 'logic of the unconscious' take their origin in the ideas of Eilhard Von Domarus, Silvano Arieti, and Ignacio Matte Blanco. Among such ideas are Von Domarus' principle that, in unconscious thinking, shared attributes imply identity (Von Domarus 1944), Matte Blanco's idea that any object is unconsciously considered as equivalent to a class of which it is a member, and his suggestion that the unconscious treats most relations as symmetrical, while some specially selected relations are treated as asymmetrical, which finally led Matte Blanco to his idea of "bi-logic" (Matte Blanco 1975; 2005). These ideas, together with the well-known general characteristics of the unconscious suggested by Freud (we will consider them in section 3) and the statement that unconscious thinking does not comply with classical logic, especially with the law of non-contradiction, constitute the essential (although not exhaustive) list of intuitions usually pursued in logical formalisations of the 'logic of the unconscious.'

Since the time Arieti and Matte Blanco put forward their projects of the 'non-Aristotelian' logic of the unconscious, contemporary logic underwent a period of enormous development. Although the idea of formalising the logic of the unconscious by non-classical symbolic logic is relatively new, we find today a variety of mathematical and logical formalisations that rely on the intuitions stated above but have expanded to the neurosciences. Thus, Matte Blanco's insights have been implemented in quantum logic and topology-based systems and laid down the scientific discipline of computational psychoanalysis (Khrennikov 2002; Lauro Grotto 2008; Battilotti 2014; Iurato 2018; Battilotti, Borozan, and Lauro Grotto 2023). Some other quantum logic formalisations have been made outside the bi-logical framework (e.g., Zizzi and Pregnolato 2012); in (Selesnick and Owen 2012), Von Domarus' principle is discussed. Besides, recently, new conceptual models for the 'logic of the unconscious' were suggested: in (Saad 2020), unconscious thinking was clarified through negationless logic; in (Saad 2022), a theory of unconscious concepts as drive-based properties was developed and applied to the analysis of unconscious processes; and in (Vasylchenko 2015), Graham Priest's logic of intentionality was applied to the analysis of

internal objects and their primary process-based dynamics. The number of studies in the area is growing, but a comprehensive review of the results has yet to be written.

Looking at the development of psychoanalysis from a broader historical perspective, we can state that in the second half of the twentieth century there was a shift in the understanding of the structuring factors of the unconscious: Freud's structural model that put the relations between id, ego, and superego in the centre of the 'inner[1] world' was replaced by a model of internal object relations. Most psychoanalysts would agree that today, "object relations theory has become a crucial starting point for psychoanalysis and its practice" (Hensel 2014, 147). The human mind, according to object relations theory, is essentially object-seeking. An '(internal) object relation' between the person and her internal objects is a cognitive-affective psychological structure derived from early interaction experiences with significant others. Internal object relations organise all our experiences and structure the unconscious mind. From the standpoint of object relations theory, much of the unconscious activity of the mind is rooted in its attempts to project internal object relations outside and reproduce them in interpersonal relationships (Auchincloss and Samberg eds. 2012, 175–178; Scharff ed. 2005).

From the perspective of object relations theory, the main focus of the logic of the unconscious shifts from the general characteristics of the unconscious to the issues that stem from the impact of subjective psychological formations – internal objects and internal object relations – on mental activity. Although the idea of reproducing *relations* by substituting external *objects* – actually, persons – instead of internal ones seems to be tempting for logical formalisation, no systematic, logical theory of internal objects and internal object relations has been proposed so far.

The task of building logical models of unconscious mental operations involving internal objects and internal object relations calls for a reconsideration and development of the ideas and intuitions lying behind the 'logic of the unconscious.' My objective in this inquiry is to outline a project on

[1] Applying the adjective 'internal' or 'inner' to some phenomena in the psychoanalytic context, we thereby indicate the intrapersonal or subjective nature of these, differentiating them from intersubjective and objective phenomena (cf. Ogden 2004, 131).

the logic of the unconscious on the grounds of object relations theory. I treat 'thinking' as a sequence of mental states involving cognitive, conative, and affective intentional attitudes. The difference between 'unconscious' and 'conscious' thinking is clarified on the basis of Freud's distinction between primary and secondary processes. To clear the ground for the articulation of philosophical and metapsychological assumptions of the logic of the unconscious from the perspective of object relations theory, I consider a number of questions. Some of them to start with are: What is the primary process? What is its place in mental life? What is the nature of internal objects and internal object relations? What is their impact on the primary process?

3. Unconscious Thinking: Primary Process and Phantasy

Freud introduced the notion of the primary process in The Interpretation of Dreams (Freud 1953) as an unconscious mental activity provided by the unconscious system (Ucs) and aimed at overcoming the tension between wish and reality through the creation of an illusion. The notion of primary process is acknowledged by most psychoanalytic authors and supported by empirical and clinical evidence (Bazan et al. 2013; Brakel 2018). According to (Freud 1957a), the most general characteristics of the primary process are replacing external with internal reality, condensation, displacement, the exemption from negation and contradictions, and timelessness (Smith 1999, 120-131; see also Rayner and Tuckett 2005, 7ff.).

In this paper, we understand 'unconscious thinking' as 'primary process mentation.' Characterising the primary process as unconscious does not imply that its mental content is necessarily repressed or sequestrated as consciously unacceptable (although it might happen to be so); instead, it is pre-reflective thinking developmentally prior to conceptual rational deliberation (Robbins 2018, 196). The primary process is unconscious in another sense: it consists of uncontrolled, non-reflective mental operations that do not require any rational justification or reality check and are ruled by defensive or adaptational 'mechanisms.' Although the primary process is not intrinsically reflective, its manifestations (arising mental contents) are often conscious. Robbins argues that the primary process is a "primordial

consciousness," and its characteristics can be found in dreaming and psychosis but also in creative and spiritual experience and mythologically driven everyday thinking.[2] Considering that the primary process is not necessarily pathological, Brakel calls this mode of thinking "a-rational" (Brakel 2009; 2018). The results of empirical research show the predominance of the primary process over rational secondary cognition in the thinking of preschool children, psychotic patients, and persons in anxiety states related to stressful situations (Brakel 2018, 179). However, even in most rational thinking, where the primary process is not predominant, it is still present in the background.

As Robbins notices (Robbins 2018, 189), there is another concept in psychoanalysis that shows a great deal of similarity with Freud's notion of the primary process: the concept of unconscious phantasy.[3] According to Melanie Klein and her followers, phantasies "underlie every mental process and accompany all mental activity" (Spillius et al. eds. 2011, 3). Phantasy has the character of unconscious thinking and unconscious psychic reality *par excellence*. Susan Isaacs argues that "reality thinking" cannot operate without "phantasy thinking," that these mental activities constantly interweave (Isaacs 1948, 93), and the latter supports the former, providing the continuity of the inner world (ibid., 80). Summarising Isaacs' approach, Thomas Odgen considers phantasy as "the process that creates meaning, and is the form in which all meanings – including feelings, defense 'mechanisms, impulses, bodily experiences, and so on – exist in unconscious mental life" (Ogden 2011, 940). Later, the philosopher Richard Wollheim (1984)

[2] A deep connection between primary process mentation and creativity has been explored in the works of Silvano Arieti (e.g., 1976; 1978).

[3] Further on, I will use the term 'phantasy' in the meaning of 'unconscious phantasy.' Although psychoanalytic theory maintains a differentiation between unconscious and conscious phantasies, emphasising links between them is no less essential (cf. Laplanche and Pontalis 1973, 316–17; Steiner 2003, 10; Bell 788). I use the spelling 'phantasy,' accepted by James Strachey in Freud's *Standard Edition*; for Strachey's argumentation in favor of this choice, see SE, vol. I, *xxiv*. Nevertheless, paying tribute to different views on the controversy in spelling, I keep intact 'fantasy' inside quotations. For the spelling controversy and the different meanings of the term, see (Steiner 2003, 54; Spillius et al. eds. 2011, 3–16; Auchincloss and Samberg eds. 2012, 85–87).

proposed an extended interpretation of phantasy as the basis of mental life. David Bell, who shares and develops Wollheim's approach, states that phantasies, supporting our investigation of reality, constitute "the 'thread' of our personhood" (Bell 2017, 793).

The concept of phantasy can be traced back to Freud's description of the case of Anna O., who was living through fairy tales in her "private theatre" (Breuer and Freud 1953; see Laplanche and Pontalis 2003, 109) – an imaginary scene of immediate wish-fulfilment, in which the person interacts with her background phantasy figures known in the Kleinian tradition as *internal objects*. Many psychoanalytic authors disagree with the expansion of phantasy into the realm of the unconscious or the inner world as a whole. However, the wish-fulfilling nature of phantasy, its role as a scene of interaction with internal figures, and its crucial influence on the structurisation of the inner world are universally recognised.[4] Thus, Joseph Sandler, writing about "unconscious fantasy presences" with whom we "have dialogues in fantasy," states:

> We can…speak of the relationship, interaction, and dialogue with an unconscious phantom presence as being intimately connected with the individual's wish fulfilments. This formulation is a much more specific one than that of simply equating the notion of an inner world with a world of unconscious fantasy, an equation that treats the whole of the inner world as being in the experiential realm... (Sandler 1990, 877).

Linda Brakel considers phantasy as the basic cognitive attitude that, unlike conscious cognitive attitudes, does not require the checking of reality (Brakel 2009). Brakel's approach, which I also take in this paper, allows us to do justice to both primary process and phantasy without conflating them. The primary process includes phantasying as its essential aspect but goes beyond it. It is a sequence of mental states involving different intentional attitudes (e.g., primary beliefs, wishes, and drives) managed by the causality of defensive and adaptational psychological 'mechanisms.' So, for example, when we are saying that a wish generates the phantasy of its fulfillment,

[4] For different psychoanalytic accounts of phantasy, see (Steiner ed. 2003).

or the combination of a phantasy with some other attitudes unconsciously generates another phantasy, we thereby appeal to primary processes.

4. Unconscious Categorisation

Now we can take a closer look at the traits of the primary process, starting from the characteristics mentioned above that were described by Freud in "The Unconscious." Brakel writes that primary processes "operate at a developmental level at which there is yet to be a stable self, capable of grasping (in any fashion) continuity-in-experience" (Brakel 2015, 132). Most fundamentally, at this level, we cannot distinguish between external reality and the "inner world of the mind … with its own dynamic laws and characteristics" (Isaacs 1952, 80). Correspondingly, primary process does not deal with any spatiotemporal objects as such, so there is no space-time and hence no past or future beyond the "unexamined present" (Brakel 2015, 131). Presumably, there is no negation either, as we start to learn what "being not the case" might mean only through a reality check. However, there is no reality check in the primary process. Furthermore, if there is no negation, there is no contradiction either, so, for example, there is nothing wrong in being dead and alive or in being a child and adult at the same time, as happens in dreams (cf. Rayner and Tuckett 2005, p. 20).

All these intrinsic features of the primary process are perfectly manifested in phantasy. In one of Brakel's examples, a person, Ms A, during her psychoanalytic session, longs for a relaxing break from work. As a result, Ms A phantasies that she is both participating in her session and, at the same time, taking a break from her analysis to enjoy a London afternoon (Brakel 2009, 113–114). Phantasy has the fortunate ability to provide an immediate illusory fulfillment of our wishes; no reality check or actual time reference is needed for such fulfillment.

The disregard for reality, the absence of time, and the toleration of contradictions are general features characterising the mental content of our phantasies. Displacement and condensation have a different nature: they can be seen as mental operations or primary process procedures rather than characteristics in the proper sense. Leaving them aside for a while, we will first explore how the unconscious mind identifies and categorises objects.

Eilhard Von Domarus pointed out that the thinking of schizophrenic patients demonstrates what he called "predicative thinking" (Von Domarus 1944, 111). Silvano Arieti summarises Von Domarus' observation in the following way: "Whereas in normal (or secondary process) thinking identity is based only upon the basis of identical subjects, in paleologic (or primary process) thinking identity is accepted upon the basis of identical predicates" (Arieti 1976, 69). Arieti provides several clinical examples confirming this feature of unconscious thinking, which he calls *identification based upon similarity*:

> A red-haired young woman in a schizophrenic post-partum psychosis developed an infection in one of her fingers. The finger's terminal phalanx was swollen and red. She told a therapist few times, 'This finger is me.' Pointing to its final phalanx, she said, 'This is my red and rotten head.' She did not mean that her finger in some way represented her; she saw it either as a duplicate of herself or, in a way that is incomprehensible to the normal person, literally herself. Another patient believed that the two men she loved in her life were actually the same person, although one lived in Mexico City and the other in New York because both of them played the guitar and both of them loved her (Arieti 1976, 68).

The ability of the unconscious mind to identify things on the basis of a common property has been extensively explored in the works of Ignacio Matte Blanco. According to Matte Blanco, the "deep unconscious" treats external spatiotemporal objects through the optics of emotively marked properties (if we describe its treatment in an intensional way) or classes of similar objects (if we put it extensionally): "the objects are, true enough, individual objects but at the same time they are not. They can at the same time be sets of objects and also general ideas or propositional functions, such as 'breastness' or 'motherhood'" (Matte Blanco 2005, 209). Furthermore, things that satisfy the same instinct or provoke the same emotion are perceived as "belonging to the same class."

> Milk, for instance, and a book are quite different things, but both are elements of the class of objects which give oral satisfaction, physical or symbolic. They are identical but only in that aspect. In the so-called deep unconscious, instead, these two objects,

> owing to the fact that they are felt or treated as elements of the same class, are not at all different; they are the same thing (Matte Blanco 2005, 224).

Similarly, in the deep unconscious "an authority is felt not simply as a substitute of the father, as is seen in classical logic; he is the father" (ibid., 39).

Brakel states that primary process thinking is "associatively based" (Brakel 2009, 7) and, in addition to the characteristics described in (Freud 1915), distinguishes "feature-based categorising by resemblance" and "categorisations by contiguity in time and/or space, and substitutions of part for whole" as hallmarks of primary thinking (Brakel 2009, 8). In another article, Brakel writes that, in the primary process, categorisations are "associatively constructed, with similarity predicated upon small, superficial, or insignificant features constituting a category" (Brakel 2018, 177).

When we speak about unconscious identification or categorisation, the choice of whether to use the term 'identification'[5] or 'categorisation' depends on whether secondary thinking is involved: where the person's unconscious mind sees two objects as being one (as both Arieti and Matte Blanco insist), the rational outsider, and the person herself, on the level of secondary thinking, would consider these two objects as falling under one concept or category (as Brakel puts it). In experientially and emotionally neutral contexts, the associations employed by primary thinking can be arbitrary, creatively constructed, and often based on (what seems to be) insignificant features, as in wit and creative art. However, as soon as emotionally significant, subjectively meaningful, painful, or traumatic mental content becomes reflected by similarities, the identifications built upon these similarities by the unconscious mind become definite, steady, and unquestionable, as Arieti's examples show.[6]

In other words, the individual unconscious mind suggests its own emotively marked reasons for classifying objects. That is, as emotional

[5] The term 'identification' can have different meanings in psychoanalytic contexts (cf. Weiss 1947). I use it here with the same meaning as Arieti when he writes about 'identification upon similarity' – namely, 'equating': to identify two objects is to state that they are identical.

[6] We will consider more examples of similarity-based identification in sections 8 and 9.

beings we are inclined to group objects according to their subjective similarities.

True, emotions are inconstant. Is there anything in the 'inner world' that could provide a solid ground for our subjectively arranged systematisation of external reality? To answer this question, we will turn to internal objects.

5. Internal Objects

In psychoanalytic contexts, 'object' refers to other persons or things, both *per se* and as they are represented in the mind (Auchincloss and Samberg eds. 2012, 170). Among objects considered in psychoanalysis are 'significant others,' primarily parents, "whose interactive impact has been internalized and continues its influence on the mind from within" (Svrakic and Zorumski 2021, 4). These previously internalised and transformed figures of our significant relationships reappear in our phantasies as 'objects' that share some characteristics of real persons, having been at the same time altered by phantasy. They might be variously called within different psychoanalytic approaches: 'introjects,' 'personifications,' 'illusory others,' and the like (Greenberg and Mitchell 1983, 11). The term 'internal object' was introduced in the early works of Melanie Klein (1984) and developed later by Ronald Fairbairn, Otto Kernberg, Joseph Sandler, Thomas Ogden, and many other authors (see Mijolla ed. 2005, 849–851; Spillius et al. eds. 2011, 40–62; and also, Mitchell 1981; Ogden 2004, 131–165; Sandler 1990; Grotstein 1997; Sandler and Sandler, 2003; Hensel 2014).

There is a difference between phenomenological and metapsychological perspectives on internal objects as phantasy figures (Spillius et al. eds. 2011, 41). Phenomenologically, internal objects are figures that inhabit our deep phantasies. Primarily, in early childhood, internal objects are the introjects of parents and other emotionally significant persons from the infant's environment. The baby feels her parents "in the concrete way in which deep unconscious phantasies are experienced" (Klein 1984b, 345). As a result, "an inner world is being built up in the child's unconscious mind, corresponding to his actual experiences of the external world and yet altered by his own phantasies and impulses" (ibid.). Later, the content of phantasies

broadens to embrace other "figures inside us, at first in simple and then in increasingly complex relations with each other" (Bell 2017, 793). During the period of growing up and adulthood, the realm of the individual's internal objects can be enriched by any encounters involving sufficient emotional or psychic intensity. The process of the 'settling down' of internal objects in the mind, called *internalisation* or *introjection*, is provided by deep phantasies.[7]

Metapsychologically, internal objects have been seen from various perspectives and interpreted within various theoretical frameworks differently. To mention just a few landmarks, they have been treated as expressions of instinctual drives (Klein 1984a), "dynamic endopsychic structures" (Fairbairn 1944), components of "object relations units" (Kernberg 1976), and unconscious suborganisations of the ego (Ogden 2004).[8] Metapsychological interpretations vary in the emphasised functions of internal objects and the assumed ways of their integration into the structure of personality. However, the variety of conceptualisations keeps intact the initial evidence of the background figures that appear in dreams, phantasies, and therapy room narratives.

> So, for example, the patient may bring experiences of being frightened of policemen. He tells the analyst he is frightened of his bank manager. He is frightened of colleagues at work, and, of course, he is frightened of the analyst. As the analytic work progresses, we can show the patient how there seems to be a frightening *internal figure* [the italics are mine – A.V.] which is externalized in the transference onto the person of the analyst (Sandler 1990a, 171).

Our primary purpose is to explore the role of background phantasy figures in unconscious thinking and to suggest a project of the logic of the

[7] Not *every* phantasy ensures the formation of internal objects. For that reason, W. R. D. Fairbairn proposed "to replace the concept of 'phantasy' by a concept of an 'inner reality' peopled by the Ego and its internal objects" (Fairbairn 1994, 559). I use the notion of 'deep phantasy' to distinguish permanent personality-structuring phantasies from a mere play of imagination that may come and go.

[8] For the analysis of different conceptualisations of internal objects, see (Ogden 2004; Grotstein 1997; Knox 1999; 131–65; Boag 2014; Brodie 2020, 125–40).

unconscious. For this purpose, we have to delineate a concept of the internal object *qua* phantasy figure that would involve as few theoretical assumptions as possible and articulate these assumptions. The more the concept is equidistant from various metapsychological frameworks, the better the resulting logical project will apply to different object relations theories.

An excellent point to start with is dictionary definitions. According to the *International Dictionary of Psychoanalysis*, "'internal object' refers to the mental representation that results from introjection, incorporation, or internalization of the relationship to an external object" (Mijolla ed. 2005, 849). In *The New Dictionary of Kleinian Thought*, we read that "the term 'internal object' means a mental and emotional image of an external object that has been taken inside the self" (Spillius et al. eds. 2011, 40) and, further, "the concept of internal object implies an object that is or has been animate and with whom there is an emotional connection. By and large, the term implies some degree of permanence or recurrence of experience of the object" (ibid., 43). Similarly, in the *APA Dictionary of Psychology*, we find that the internal object is "an image or representation of a person (particularly someone significant to the individual, such as a parent) that is experienced as an internalized 'presence' within the mind" (VandenBos ed. 2015, 553).

In all these definitions, first, the internal object is considered as a 'representation' or 'image.' Second, an internal object, being more than a creative play of imagination, is characterised by permanence, recurrence, or 'presence' in the mind. Third, the internal object is invested with emotional content: it is a "mental representation of another person that one invests with drives and affects" (Svrakic and Zorumski 2021, 4).

6. Imagos and Imagery

In this section, we will try to find out what kind of 'representation' or 'image' the definitions of the internal object involve.

Klein, in her earliest writings, before coining the term 'internal object,' used the term 'imago' to refer to phantasy figures that substitute parents in the child's mind (see Perlow 1995, 27–31). Isaacs employs the contrast between 'imago' and 'image' when she explains internalisation dynamics.

> The earliest phantasies are built mainly upon oral impulses, bound up with taste, smell, touch (of the lips and mouth), kinesthetic, visceral, and other somatic sensations; these are at first more closely linked with the experience of 'taking things in' (sucking and swallowing) than with anything else. The visual elements are relatively small (Isaacs 1948, 92).

Later, "when the visual elements in perception…begin to predominate over the somatic, … making clearer the distinction between the inner and the outer worlds, the concrete bodily elements in the total experience of perceiving (and phantasying) largely undergo repression" (ibid., 92–93). Finally, the visual elements become 'images,' "representations 'in mind' (but not, consciously, incorporations in the body) of external objects recognized to be such. It is 'realized' that the objects are outside the mind, but their images are 'in the mind'" (ibid., 93).

So, *imago* can be considered as a proto-image that is yet to be internalised; unlike the image, it is not 'within,' not yet *internal.* As unconscious image, "'imago' … usually refers to a person or part of a person" (ibid.). Further,

> 'imago' includes all the somatic and emotional elements in the subject's relation to the imaged person, the bodily links in unconscious phantasy with the id, the phantasy of incorporation which underlies the process of introjection; whereas in the 'image,' the somatic and much of the emotional elements are largely repressed (Isaacs 1948, 93).

It is plausible that the internalised image (internal object) that originates from an initial imago preserves the repressed somatic and emotional content based on its hereditary link with the imago. Thus, Maria Torok makes a distinction between an internal object as "the fantasmic pole of the introjective process" and the imago as "precisely all that resisted introjection" (Torok 1968, 121). If so, we can assume that what the subject experiences as a (usually visual) image of a phantasy figure is, in fact, a more complex, even if repressed, representation that involves different senses, such as auditory, tactile, gustatory, and olfactory. The repressed emotional content towards the figure might manifest in phantasies or dreams; it can also make its way to reality through occasional slips of the tongue or involuntary behavioural dispositions.

The type of representation that allows various sensory modalities and thus might be relevant for representing phantasy figures is known as *imagery* (also 'mental imagery'). In the *APA Dictionary of Psychology*, imagery is defined as the "cognitive generation of sensory input from the five senses, individually or collectively, which is recalled from experience or self-generated in a nonexperienced form" (VandenBos ed. 2015, 524). Bence Nanay defines imagery as "perceptual representation not triggered directly by sensory input. This leaves open the question what it is triggered by" (Nanay 2021, 1.1).

"If you close your eyes and visualize an apple, what you experience is mental imagery – visual imagery. But mental imagery is far more pervasive in our mental life than just visualizing. It happens in all sense modalities" (Nanay 2021). "Just as perception can be visual, auditory, olfactory, tactile, gustatory, etc., the same goes for mental imagery" (ibid., 1.2). Mental imagery is often involuntary and ensured by early cortical processes; it is sensitive to the allocation of attention.

Furthermore, mental imagery, just as perception, can be unconscious as well as conscious (Nanay 2021a). Typical examples of involuntary mental imagery are flashbacks, earworms, hearing voices, phantosmia, and visual hallucinations. In each of these cases, there is a change in the cortex subjectively interpreted as image, sound, or smell; however, there is nothing on the retina, eardrum, and smell receptors. What takes place in the cortex is not caused by any external stimuli.

Phantasies involve imagery: in the process of phantasying, consciously or unconsciously, we subjectively sense the figures of our phantasies as psychically present. Edoardo Weiss, describing the phenomenon of the "psychic presence" of such figures, writes that the presence can reveal itself not only by a visual image but also "in an acoustical way," for example, as the voice of a crying mother (Weiss 1939, 181). The term 'image' applies in psychoanalytic contexts "to conscious or unconscious experiential data arising from all the sensory modalities, including proprioceptive or kinesthetic experiences" (Sandler 1990a, 166).

That is, phantasy figures look like persons and sound like persons; they are like persons by touch and move us emotionally in the same way as persons do.

If so, why are internal objects defined as representations and not as *representables*?

7. Internal Objects as Representations

Meir Perlow, the author of an influential work on mental objects (Perlow 1995), states that the concept of representation in psychoanalysis differs from that in cognitive psychology and points out two aspects of the difference. First, representations in psychoanalysis have emotional-cognitive rather than merely cognitive character; second, psychoanalysis is concerned more with the influences of drives and wishes on a representation than with its adequacy in representing reality (Perlow 1995, 2). Perlow defines the psychoanalytically relevant concept of representation as an anticipatory set that organises experience:

> Thus 'object representations' were not considered to be only specific 'images' of objects revived in response to the awakening of various wishes …, but also the amalgamation of all the experiences the individual has of his objects, including his actual interactions with them and their emotional meanings, as well as the distortions of realistic aspects under the influence of drives and phantasies. As such, a mental representation of an object refers to a 'schema' which, on the basis of past experience (not necessarily realistic), organizes present experience and provides a context for both present perceptions and for the recall of past memories" (Perlow 1995, 120).

According to Perlow, representation is just a theoretical construct not subject to first-person experiences, while internal objects as special objectual phantasies are experiences, 'psychic presences' (ibid., 120–127). So, on this account, internal objects are *not* representations.

For Joseph Sandler, on the contrary, internal objects *are* representations and not first-person experiences. Sandler's theory of internal objects integrates the Kleinian treatment of internal objects as phantasy figures into non-Kleinian mainstream psychoanalysis. We will consider Sandler's approach in more detail.

Sandler states that the term 'representation' "can … be a source of confusion, in that it relates to subjective experience on the one hand and to what can be called a nonexperiential organization on the other" (Sandler 1990a, 166). He emphasises the significance of the differentiation "between the experiential content of a mental representation – the perceptual and ideational content – and the structural organization behind that content, an organization that lies outside the realm of conscious or unconscious experience" (Sandler 1990, 869). In the centre of Sandler's approach is the distinction between the experiential and nonexperiential "representational realms."

> The experiential realm can be regarded as the world of subjective experience, the phenomenal world projected, as it were, onto different aspects of an experiential screen. It is the world of perceptual images, of sensations, of memory content and of the content of visualization and imagination; it is also the world of feelings.…The nonexperiential realm is the world of biological and psychological structures and apparatuses in the sense of enduring organizations. It is a world of functions, of plans and schemata, of forces and energies and of the tools of intrapsychic adaptation (Sandler 1990a, 165).

Accordingly, internal objects can be considered as 'representations' in two distinct senses of the term: as perceptual images in various sensory modalities and as nonexperiential "enduring organizations" "structured by subjective experience" and "capable in turn of affecting perception, memory, and fantasy" (Sandler 1990, 870). Sandler criticises Kleinians for the lack of distinction "between the processes involved in the creation of a perceptual image of the object…and the setting up of the object as an internal authority figure" (ibid., 863). He claims that the term 'internal object' in its most appropriate meaning applies not to an image but to the background "enduring organization" or "structure" shaping phantasy images (ibid., 871).

In Sandler's view, there is a two-way connection between perceptual images appearing in phantasy and thought and internal objects as enduring organisations lying behind them. On the one hand, internal objects are built up on the basis of the perceptual and phantasy experience of the person. On the other hand, they are the source of new phantasy images – "internal

'presences' with which the person constantly unconsciously interacts" (ibid., 859). Sandler describes this two-way connection as follows:

> So, I want to suggest that the world of internal objects in the nonexperiential realm be considered as giving rise to a fantasy and thought, to a (largely unconscious) phantom world in which we live at the same time as we live in the real world. This inner world is a world of unconscious ghost object images which we have constructed during the course of development, objects in unconscious fantasy whose presence is gratifying, wish-fulfilling, and reassuring. While we can conceive of the internal objects as relatively unchanging structures, the figures in this fantasy world, derivatives of the internal objects, do change as they are updated and modified… (Sandler 1990, 877–878).

The connection between an internal object and its experiential representations – momentary images or 'presences' – is further clarified by Sandler's following clinical example:

> A patient brings a report that he was afraid when a policeman stopped his car, sure he had done something seriously wrong, but the policeman only wanted to tell him a rear light needed repair. The patient tells of receiving a letter from the tax authorities and of his panic before opening it at the thought of the tax evasions he might be accused of and penalties he would face. He brings several similar reports, and the analyst is able to show the patient how afraid he is of being accused of some awful misdeed by the analyst. From material such as this, with its obvious recurring theme, the analyst may eventually construct a notion of a threatening internal object which is externalized in a variety of situations … The internal object, as it has been conceived of here, is only evident in the shape of its derivatives. What is perceived by analyst and patient is referred to the hypothetical internal object in its relation to the patient's own self (Sandler 1990, 875).

The patient's reports contain descriptions of concepts and images – a policeman, tax authorities, the analyst – that not only *refer* to a background figure but also *represent* it. Sandler writes that 'internal objects' should be regarded as either theoretical or clinical constructs, "organizing concepts

employed in the conceptualization of the patient's inner psychic structure" (Sandler 1990a, 170). So, they are not first-person representations or, as Sandler says, not experiential representations. Instead, 'internal object' is a third person construct that refers to what is experienced from the first-person perspective as a background phantasy figure. That is, both theoretically and clinically, internal objects *are* background phantasy figures.

Sandler's nonexperiential "world of psychological structures and apparatuses" is the theoretical world to which psychoanalytic terms refer and in which they require their meaning. From the semantical point of view, this world should be characterised not as a separate domain of representations but as the theoretical realm of abstract objects, their properties, functions, relations, and causal processes, just like various other theoretical realms, for example, those of mathematical analysis, theoretical chemistry, or quantum physics. Within this perspective, internal objects are abstract objects; we will further develop this approach in section 10.

8. Object Relations

Object relations theory is one of the major schools in contemporary psychoanalysis (Scharff ed. 2005; Auchincloss and Samberg eds. 2012, 175–178). In this section, for the sake of clarity and completeness, we will explicitly state its main ideas. Due to multiple differences and variations, it is more relevant to speak about object relations theories rather than theory; however, we will focus only on the common ground shared by different approaches.

It is within the framework of object relations that internal objects obtain their psychological significance. An *object relation*, also called *internal object relation* or *internal object relationship*, is an intrapsychic structure consisting of three parts: a self-component, an object-component, and a representation of an affectively charged interaction between them (Auchincloss and Samberg eds. 2012, 175). For example, an internal object 'mother' is always related to an internal self-referring component, say 'child,' as its counterpart. We apply the term 'internal object' to refer to either the self or the object component of the internalised relations. Internalised object relations supply the person with potential relationship roles, the accompanying dispositions and expectations towards others.

Contemporary studies in early developmental psychology (e.g., Chambers 2017; Holmes and Slade 2019) confirm John Bowlby's seminal idea that infants have "an inbuilt need to be in touch with and to cling to a human being" (Bowlby 1958, 350). Our need to be in attachment relationships is independent of the needs for food and warmth, and as fundamental as they are. This lifelong need creates the ground on which all our internal objects and internal object relations are formed and developed. As Sandler writes, "we can speak of attachment between a person and a fantasy object just as we can speak of attachment to the real external objects in the person's life" (Sandler 2003, 16).

The main points of object relations theory can be formulated, according to (Auchincloss and Samberg eds. 2012), in the form of the following four postulates.

1) All psychological experience, from the most fleeting fantasy to the most stable structure, is organized by object relations. In other words, object relations are the basic unit of experience.
2) The human mind is object-seeking from birth; the basic motivation for object-seeking is not reducible to any other motivational force (as, for example, drive in Freud's theory).
3) Internalized object relations are built up in the course of development through the interaction of innate factors (such as inborn affect dispositions and cognitive equipment) and relationships with others (primarily caregivers).
4) Interpersonal relationships reflect internalized object relations; psychopathology, especially serious psychopathology such as psychosis, and borderline and narcissistic personality disorders, is best conceptualized in terms of object relations (Auchincloss and Samberg eds. 2012, 175).

As a result of the development of object relations theory, the inner world is often seen in psychoanalytic metapsychology as being structured by previously internalised patterns of the person's relationships with significant others. Thus, Thomas Ogden writes about "the development of a radically revised psychoanalytic theory of thinking" consisting of "the replacement of Freud's structural model with a model of an inner world structured by phantasied internal object relationships" (Ogden 2011, 926). Internal object relations constitute the quintessence of our previous experience of

interactions and thus create the optics through which we perceive the external world. This optics is being formed and transformed by different primary mental operations and processes considered in object relations theory: introjection, projection, identification, projective identification, transference, condensation, displacement, splitting, and the like. All these operations contribute to the formation, development, transformation, or reactivation of the patterns of internal object relationships involved (see, e.g., Ogden 2004; Sandler ed. 1988). To get an idea of the dynamics of the development of internal objects, we will consider closer transference, displacement, and condensation.

Transference is "the externalization of an internal object relationship" (Ogden 2004, 151), which is the opposite of internalization. As mentioned above, the internal object relationship comprises the self-component and the object component. Correspondingly, "transference can take one of two forms, depending on whether it is the role of the object or that of the self in the internal object relationship that is assigned to another person in the externalization process" (ibid.). As long as 'internal object' can refer to either the self or the object component of the internalised relations, the *externalisation* (or *projection*) of an internal object can refer to either form of transference.

An example of the externalisation of the self-component can be found in Freud's paper on Leonardo da Vinci, in which Freud analyses the formation of Leonardo's phantasy of love: "The boy represses his love for his mother: puts himself in her place, identifies himself with her, and takes his own person as a model in whose likeness he chooses the new objects of his love" (Freud 1957, 100). Commenting on Freud's analysis, Bell writes that Leonardo "needs to sustain in his mind a particular scene/phantasy where he is being loved by his mother, and to achieve this he distributes different elements of this phantasy between two figures: himself and his lover" (Bell 2017, 788).

Displacement is a shift of ideas or feelings from one object to another. Eric Rayner describes a patient who was afraid of spiders:

> …it was fairly easy to discover that her mother and the class of spiders had several qualities in common, such as activity akin or identical to weaving webs that entrap. Spiders in themselves were known to be harmless, but, as they had something in common

> with her mother, they took on other maternal attributes, and then being afraid. A defense of denial may then come into play so that the idea or feeling is not consciously recognized as belonging to the primary object but only to the secondary one – the patient was thus not afraid of her mother but of spiders instead (Rayner 1995, 42–43).

Finally, *condensation* is a conflation of features of different external objects in one internal object. Rayner gives the following example:

> Before a holiday, a patient dreamt that she was going down a ski slope with her instructor. Part of the way down he stopped but she continued the rest of the way easily by herself. The session then showed how she had experienced parting from 'givers of instruction' at different times of her life. The ski-instructor could be recognized as a condensed timeless image of her mother, father, elder brother, and her analyst, as well as the knowledgeable side of herself (Rayner 1995, 42).

It is worthwhile to note that, in each of the examples of unconscious mental operations considered above, different objects are seen in phantasy as identical upon similarity: they are similar in *emotionally charged* properties of internal objects. In Leonardo's case, described by Freud, the identifying property is 'looking like young Leonardo,' where 'young Leonardo' is an internal object. In the example of displacement, the reason for unconscious identification is 'weaving webs that entrap'; in condensation, it is the property of 'being a giver of instructions.'

So, coming back to the question of what can provide a ground for subjective systematisation of external reality, raised in section 4, we can suppose that phantasied internal object relations will do.

9. Transference and Distinctive Similarities

In this section, we will consider two examples illustrating the role of identification upon similarity in transference.

Transference gives humans a chance to connect their inner world with reality. It occurs because "there is a constant need in every individual to

externalize his or her internal objects (the introjects), to anchor the inner world as far as possible in external reality" (Sandler 2003, 19–20). The unconscious mind chooses an object for externalisation due to an emotionally charged property shared (or perceived as such) by the internal and external object. To put it otherwise, some perceived properties of real persons can activate in us phantasies of our internal objects. As Sandler puts it, "transference elements are present to a varying degree in all relationships and are often determined by some characteristic of the other person who has an attribute of an important figure of the introject" (ibid., 19). We will call such emotionally charged properties capable of provoking transference *distinctive similarities*.

The following case description – *Walter and Sandra* – is borrowed from Robert Kraut:

> Sandra and Walter have been living together for some time. He tells her that he loves her. He undoubtedly exhibits the kind of behavior toward her that we are inclined, given other details of Walter's global psychology, to call love behavior. Sandra is upset. One evening she tells him: 'It isn't really me you love.' Walter is confused. 'But you're the only one I share my joys and sorrows with, the one I most enjoy talking to, etc. And you're the only one I sleep with.' Sandra says: "I know I'm on the receiving end of your love behavior. But that's not enough. I want to be on the receiving end of your love. And I'm not' (Kraut 1986, 413–414).

Sandra's predicament can be explained by the account that we can feel love, hatred, fear, desire, or repulsion not only towards persons and their properties but also towards our internal objects. Thus, we can assume that something in Sandra, namely, a distinctive similarity between Sandra and Walter's internal object initially related to his mother – say, her hairstyle or manner of walking – has once provoked Walter to project his maternal figure onto Sandra. We can also imagine that Walter experienced a deficiency in his mother's care for him as a child. Now, identifying Sandra with his maternal figure, he externalises not only the figure itself but also his wish for maternal care, trying to make up for the deficiency in this way and being unaware of this aspect of his feelings towards Sandra.

Once the transference has happened, in Walter's phantasy, the distinctive similarity merges with his desperate unconscious striving for maternal care. Walter might become rigidly attached (let us imagine) to Sandra's hairstyle: of course, he wants it to remain constant, for this constancy gives him a (perhaps illusory) promise that his strivings after care will eventually be satisfied. Walter's attachment is directed primarily toward his internal love object and only secondarily, due to displacement, toward Sandra. Furthermore, Sandra might perfectly experience what happens, as the actualisation of transference in relationships is often accompanied by "subtle attempts to manipulate others" (Sandler 2003, 19).

Walter's example also shows that internal objects are not fixed once and forever. Instead, each internal object results from a series of internalisations, deep fixations, and externalisations, usually occurring on the ground of an internalised pattern of significant relationships. When Sandra becomes a part of Walter's internal love object, the object acquires her properties: for example, her femininity, her manner of movement, hairstyle, the blue color of her eyes, the touching way she whispers and raises her voice, her unique manner of giving irrelevant answers or keeping silent when she does not want to answer. In addition, this very object is endowed in Walter's phantasy with his mother's loving care, even though Walter, in real life, had never experienced such care. Furthermore, one should not be surprised if, due to condensation, the object contains Walter's aunt's charming way of giving in to pressure and even a unique smile typical for Marie-Louise, Walter's first teacher of French.

Marvel and his hamster. Marvel is a man who once, as a little boy, lived on his father's neglected farm in Illinois, far from the big cities. They mostly lived there together with his father. His father cherished Marvel and, at the same time, controlled him. This combination of tenderness and control created a feeling of safety for Walter. On Sundays, his father used to take Marvel to the nearest town for ice cream – of course, only if little Marvel earned such a trip with his behavior. Marvel was happy on his father's farm in Illinois.

Marvel is now over 40, and he is lonely. Looking in the mirror, he sees more and more features of his father. Marvel has tenderness inside him and carries it within himself. He also has a hope, or rather, a wishful phantasy, that one day he will have someone to care for.

One day, Marvel goes to the pet store and buys a smooth-haired hamster there. The hamster has chubby cheeks – exactly like little Marvel in his childhood photos. He also buys a cage for the hamster. Since this day, Marvel's life has changed because now he has someone to cherish and control. His phantasy is fulfilled; Marvel internally returns to his father's farm.

In this example, Marvel not just accidentally externalises his internal object but acts out of his phantasy to externalise his internal father-son relations upon a distinctive similarity between 'little Marvel' and the hamster (chubby cheeks).

10. Towards the Logic of Object Relations

Now, we are in a position to articulate the main assumptions of the logic of the unconscious from the perspective of object relations theory. The assumptions presented below summarise our analysis of psychoanalytic notions; they are made from the logical point of view and thus involve logical and semantic idealisations where necessary.

The assumption of particularity. Internal objects are neither images nor representations (*pace* dictionary definitions, cf. sections 5-7); they *are abstract particulars of their own kind*. An internal object as a *representable* should be distinguished from its representations, just as we distinguish a fictional object from its descriptions or a natural number from its definitions.

The same internal object can have different representations. For example, the images of two different persons can represent the same object, as in Arieti's example of the patient who believed that two men are the same person (section 4).

Each internal object, being an abstract particular, has its properties, including intentional properties perceived or conceived by the first person. For example, 'weaving webs that entrap' and 'being a giver of instructions' are conceptual first-person properties of the corresponding internal objects in Rayner's examples (section 8).

The assumption of intentionality. Our emotive *objectual attitudes*, such as love, hatred, fear, anger, envy, jealousy, or attachment, can be directed towards internal as well as external objects. Similarly, our cognitive, conative, and emotive *propositional attitudes* can be directed towards states of

affairs involving properties of internal objects and internal object relations. Some of our propositional attitudes are unconscious, that is, are not based on reflection, and, as mental states, can occur out of the first-person awareness. Internal objects possess their perceptual and ideational properties *in phantasy*, which makes them background phantasy figures.

The assumption of object relations. Phantasied relations between the self and its internal objects are basic psychological structures that supply the person with potential relationship roles, the accompanying dispositions, and expectations towards others.

The assumption of the primary process. The mental process (or thinking) is a sequence of mental states involving intentional attitudes. The primary process (section 3) is a particular kind of mental process that consists only of uncontrolled mental operations, such as internalisation, transference, identification, and the like (section 8), managed by the causality induced by defensive and adaptational psychological 'mechanisms.'

The assumption of similarity-based transference. Transference is a primary mental operation grounded in the object-seeking nature of the mind and its need to externalise internal object relations. It consists of creating the phantasy of identity between the internal object and a real person or thing. *Transference is caused by a distinctive similarity* – an emotionally charged property shared by the internal and external object (section 9). Distinctive similarities are also involved in other primary mental operations, particularly displacement and condensation.

These five assumptions create a ground on which the logic of the unconscious can be built as a logic of object relations. The necessary logical tools of such logic must include intentional modalities, an implication that formalises causality, and equality. Other requirements to the logic of object relations include providing an account for the cases of replacing perception with phantasy, formalising atemporality and paraconsistency regarding unconscious contents, modelling transference, displacement, condensation, and other primary mental operations, and explicating the impact of identification upon similarity on perception and other intentional attitudes, such as beliefs.

The logic of object relations should be easily applicable to psychotherapeutic and psychiatric cases and to other mental phenomena that involve unconscious intentionality.

11. Conclusions and Further Issues to Explore

In this article, relying on the study of the notion of the internal object in psychoanalysis and considering the main psychoanalytic approaches to unconscious thinking and object relations theory, I suggested a logically grounded interpretation of internal objects as abstract particulars. The logic of object relations built on the ground of the assumptions suggested in section 10 should include in its syntax the language of first-order logic with equality, operators of intentional attitudes, and so-called 'triggering' implication modelling causality. A possible world semantics of such logic can be built on the basis of the semantics of intentionality introduced by Graham Priest (2005), taking into account the theory of primary psychological attitudes developed by Linda Brakel (2009).

In my earlier paper, I proposed a preliminary sketch of the required logic type (Vasylchenko 2015). However, the logic of object relations is yet to be developed more systematically and axiomatised on the basis of the research undertaken in this article. Presumably, the logic of object relations will include three types of axioms: axioms of intentionality, axioms of the universal properties of internal objects and internal object relations, and axioms of causal primary process regularities.

Another direction for future research involves further clarification of the ontological status of internal objects. Internal objects are a kind of 'non-existents' – objects lacking actual spatiotemporal existence – on a par with fictional objects, abstract, past and future objects, and causally counterfactual objects. The universal properties of internal objects as a particular sort of non-existents are yet to be explicated. I hypothesise that, unlike possible spatiotemporal objects, internal objects are *possible* abstract particulars, and unlike objective mathematical or intersubjective fictional objects, internal objects are *subjective* possible abstract particulars.[9]

[9] This paper was written as part of the grant project No. 20-18675S "The nature of logical forms and modern logic" supported by the Czech Science Foundation and coordinated by the Institute of Philosophy of the Czech Academy of Sciences in Prague.

Acknowledgements

I would like to thank Jarda Peregrin, Vít Punčocář, Amit Saad and Vladimír Svoboda for valuable comments and fruitful discussions on earlier drafts, and Greg Evans for language editing of this paper.

References

Arieti, S. 1955. "Some Aspects of Language in Schizophrenia." In *On Expressive Language: Papers Presented at the Clark University Conference on Expressive Language Behavior*, edited by H. Werner, 53–67. Worcester, Massachusetts: Clark University Press. https://doi.org/10.2307/411370

Arieti, S. 1976. *Creativity. The Magic Synthesis*. New York: Basic Books.

Arieti, S. 1978. "From Primary Process to Creativity." *J Creat*, 12(4): 225–46. https://doi.org/10.1002/j.2162-6057.1978.tb00171.x

Auchincloss, E. L. and E. Samberg, eds. 2012. *Psychoanalytic Terms and Concepts*. New Haven, Connecticut: Yale University Press.

Battilotti, G. 2014. "Symmetry vs. Duality in Logic: An Interpretation of Bi-Logic to Model Cognitive Processes Beyond Inference." *Int J Cognit Informat Nat Intell*, 8(4): 83–97. https://doi.org/10.4018/ijcini.2014100105

Battilotti, G., Borozan M., and Lauro Grotto, R. 2023. "The Modal Components of Judgements in a Quantum Model of Psychoanalytic Theory." *Entropy*, 25: 1057. https://doi.org/10.3390/e25071057

Bazan, A., D. Kim Van, L. De Kock, L. A. W. Brakel, F. Geerardyn, and H. Shevrin, H. 2013. "Empirical evidence for Freud's theory of primary process mentation in acute psychosis." *Psychoanal Psychol*, 30: 57–74. https://doi.org/10.1037/A0027139

Bebee, H., Effingham, N., Goff, P. 2011. *Metaphysics: The Key Concepts*. London: Routledge. https://doi.org/10.4324/9780203835258

Bell, D. 2017. "Unconscious Phantasy: Some Historical and Conceptual Dimensions." *Int J Psychoanal*, 98(3): 785–98. https://doi.org/10.1111/1745-8315.12586

Berto, F. and M. Plebani. 2015. *Ontology and Metaontology: A Contemporary Guide*. London: Bloomsbury. https://doi.org/10.5040/9781474216166

Berto, F., F. Kroon, A. and Voltolini. 2014. "Metaontology: Introduction." *The Monist* (special issue), 97(4): 423–29. https://doi.org/10.1093/monist/97.4.423

Boag, S. 2014. "Ego, Drives, and the Dynamics of Internal Objects." *Frontiers in Psychoanalysis and Neuropsychoanalysis*, 5: 1–13. https://doi.org/10.3389/fpsyg.2014.00666

Bowlby, J. 1958. "The Nature of the Child's Tie to His Mother." *Int J Psycho-Anal*, 39: 350–73.

Brakel, L. A. W. 2009. *Philosophy, Psychoanalysis, and the a-Rational Mind.* Oxford: Oxford University Press. https://doi.org/10.1093/med/9780199551255.001.0001

Brakel, L. A. W. 2018. "The Primary Process: Bridges to Interdisciplinary Studies of Mind." *Psychoanal Inq*, 38(3): 177–85. https://doi.org/10.1080/07351690.2018.1430964

Brentano, F. 1995 [1874]. *Psychology from an Empirical Standpoint*, edited by O. Kraus, translated by A. C. Rancurello, D. B. Terrell and L. L. McAlister. London: Routledge.

Breuer, J. and Freud, S. 1953 [1895]. *Studies on Hysteria*. SE II: 1–335.

Brodie, B. 2020. *Object Relations and Intersubjective Theories in the Practice of Psychotherapy*. London: Routledge. https://doi.org/10.4324/9780367855697

Carnap, R. (1956). *Meaning and Necessity: A Study in Semantics and Modal Logic*. Chicago: University of Chicago Press.

Chambers, J. 2017. "The Neurobiology of Attachment: From Infancy to Clinical Outcomes." *Psychodyn Psychiatry*, 45(4): 542–63. https://doi.org/10.1521/pdps.2017.45.4.542

Fairbairn, W. R. D. 1944. "Endopsychic Structure Considered in Terms of Object-Relationships." *The Int J Psychoanal*, 25: 70–93.

Fairbairn, W. R. D. 1994 [1943]. "Phantasy and Inner Reality." In *From Instinct to Self*, edited by D. E. Scharff and E. F. Birtles, vol. II, 557–60. Northvale, New Jersey: Jason Aronson.

Freud, S. [SE] 1953-1964. *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, edited by J. Strachey, 24 vols. London: Hogarth Press and the Institute of Psychoanalysis.

Freud, S. 1953 [1899]. *The Interpretation of Dreams*. In Freud, S., SE IV–V: ix–627.

Freud, S. 1957 [1910]. "Leonardo Da Vinci and a Memory of His Childhood." In Freud, S., SE XI: 57–138.

Freud, S. 1957a [1915]. "The Unconscious." In Freud, S., SE XIV: 159–90.

Greenberg, J. and Mitchell, S. 1983. *Object Relations in Psychoanalytic Theory*. Cambridge, Massachusetts: Harvard University Press. https://doi.org/10.2307/j.ctvjk2xv6

Grotstein, J. S. 1997. "'Internal Objects' or 'Chimerical Monsters'?: The Demonic 'Third Forms' of the Internal World." *J Anal Psychol*, 42(1): 47–80. https://doi.org/10.1111/j.1465-5922.1997.00047.x

Hensel, B. F. 2014. "On The Origin of Internal Objects in the Works of Fairbairn and Klein and the Possible Therapeutic Consequences." In *Fairbairn and the*

Object Relations Tradition, edited by G. S. Clarke and D. E. Scharff, 147–60. London: Karnac. https://doi.org/10.4324/9780429474538-13

Holmes, J. and Slade, A. 2019. "The Neuroscience of Attachment: Implications for Psychological Therapies." *Br J Psychiatry*, 214(6): 318–19. https://doi.org/10.1192/bjp.2019.7

Isaacs, S. 1948. "The Nature and Function of Phantasy." *Int J Psychoanal*, 29: 73–97.

Iurato, G. 2018. *Computational Psychoanalysis and Formal Bi-Logic Frameworks*. Hershey: IGI Global. https://doi.org/10.4018/978-1-5225-4128-8

Kernberg, O. 1976. *Object-Relations Theory and Clinical Psychoanalysis*. New York: Jason Aronson.

Khrennikov, A. 2002. *Classical and Quantum Mental Models and Freud's Theory of Unconscious/Conscious Mind*. Växjö: Växjö University Press.

Klein, M. 1984. *Love, Guilt and Reparation and Other Works 1921–1945*. London: Virago Press Ltd.

Klein, M. 1984a [1930]. "The Importance of Symbol-Formation in the Development of the Ego." In Klein, M., 1984, 219–32.

Klein, M. 1984b [1940]. "Mourning and Its Relation to Manic-Depressive States." In Klein, M., 1984, 344–69.

Knox, J. 1999. "The Relevance of Attachment Theory to a Contemporary Jungian View of the Internal World: Internal Working Models, Implicit Memory and Internal objects." *J Anal Psychol*, 44(4): 511–30. https://doi.org/10.1111/1465-5922.00117

Kraut, R. 1986. "Love De Re." *Midwest Stud Philos*, 10: 413–30. https://doi.org/10.1111/j.1475-4975.1987.tb00549.x

Laplanche, J. and J.-B. Pontalis. 1973. *The Language of Psychoanalysis*. London: Hogarth Press & The Institute of Psychoanalysis.

Laplanche, J. and J.-B. Pontalis. 2003 [1968]. "Fantasy and the Origins of Sexuality." In Steiner, R., 2003, 107–43. https://doi.org/10.4324/9780429484469-5

Lauro Grotto, R. 2008. "The Unconscious as an Ultrametric Set." *The American Imago*, 64(4): 52–62. https://doi.org/10.1353/aim.2008.0009

Matte Blanco, I. 1975. *The Unconscious as Infinite Sets: An Essay in Bi-Logic*. London: Duckworth. https://doi.org/10.4324/9780429483592

Matte Blanco, I. 2005 [1988]. *Thinking, Feeling, and Being: Clinical Reflections on the Fundamental Antinomy of Human Beings and World*. London: Routledge. https://doi.org/10.4324/9780203358849

Meinong, A. 1904. "Über Gegenstandstheorie." In *Untersuchungen zur Gegenstandstheorie und Psychologie*, edited by A. Meinong, 1–51. Leipzig: J. A. Barth.

Mijolla, Alain de, ed. 2005. *International Dictionary of Psychoanalysis*, vols. 1-4. New York: Thomson Gale.

Mitchell, S. 1981. "The Origin and Nature of the 'Object' in the Theories of Klein and Fairbairn." *Contemp Psychoanal*, 17: 374–98. https://doi.org/10.1080/00107530.1981.10745670

Nanay, B. 2021. "Mental Imagery." *The Stanford Encyclopedia of Philosophy* (Winter 2021 Edition), edited by E. N. Zalta. https://plato.stanford.edu/archives/win2021/entries/mental-imagery/

Nanay, B. 2021a. "Unconscious Mental Imagery." *Phil Trans R Soc*, B 376: 20190689. https://doi.org/10.1098/rstb.2019.0689

Ogden, T. H. 2004 [1986]. *The Matrix of the Mind: Object Relations and the Psychoanalytic Dialogue*. New York: Rowman & Littlefield Publishers. https://doi.org/10.4324/9780429482403

Ogden, T. H. 2011. "Reading Susan Isaacs: Toward a Radically Revised Theory of Thinking." *Int J Psychoanal*, 92: 925–42. https://doi.org/10.1111/j.1745-8315.2011.00413.x

Perlow, M. 1995. *Understanding Mental Objects*. London and New York: Routledge. https://doi.org/10.4324/9780203360187

Priest, G. 2005. *The Logic and Metaphysics of Intentionality*. Oxford: Oxford University Press.

Quine, W. V. O. 1948. "On What There Is." *The Review of Metaphysics*, 2(5): 21–38.

Rayner, E. 1995. *Unconscious Logic: An Introduction to Matte Blanco's Bi-Logic and its Uses*. London: Routledge. https://doi.org/10.4324/9780203360262

Rayner, E. and Tuckett, D. 2005. "An Introduction to Matte Blanco's Reformulation of the Freudian Unconscious and His Conceptualization of the Internal World." In Matte Blanco, I., 2005, 3–34. https://doi.org/10.4324/9780203358849-6

Reicher, M. 2022. "Nonexistent Objects." *The Stanford Encyclopedia of Philosophy* (Winter 2022 Edition), Edward N. Zalta & Uri Nodelman (eds.). https://plato.stanford.edu/archives/win2022/entries/nonexistent-objects/

Robbins, M. 2018. "The Primary Process: Freud's Profound Yet Neglected Contribution to the Psychology of Consciousness." *Psychoanal Inq*, 38(3): 186–97. https://doi.org/10.1080/07351690.2018.1430965

Russell, B. 1905. "Untersuchungen zur Gegenstandstheorie und Psychologie. Herausgegeben von A. Meinong." *Mind*, 14: 530–38.

Saad, A. 2020. "On the Logic of the Unconscious." *Int J Psychoanal*, 101(2): 239–56. https://doi.org/10.1080/00207578.2020.1726713

Saad, A. 2022. "On the Logic of Unconscious Objects." *Int J Psychoanal*, 103(4): 581–600. https://doi.org/10.1080/00207578.2022.2028551

Sandler, J. 1990. "On internal Object Relations." *J Amer Psychoanal Assn*, 38: 859–80. https://doi.org/10.1177/000306519003800401

Sandler, J. 1990a. "On the Structure of Internal Objects and Internal Object Relationships." *Psychoanal Inq*, 10(2): 163-81. https://doi.org/10.1080/07351699009533805

Sandler, J. 2003. "On Attachment to Internal Objects." *Psychoanal Inq*, 23(1): 12–26. https://doi.org/10.1080/07351692309349024

Sandler, J. and A.-M. Sandler. 2003 [1998]. *Internal Objects Revisited*. London: Karnac Books. https://doi.org/10.4324/9780429476099

Sandler, J., ed. 1988. *Projection, Identification, Projective Identification*. London: Karnac Books. https://doi.org/10.4324/9780429478567

Scharff, D. E., ed. 2005. *Object Relations Theory and Practice: An Introduction*, selected and integrated by D. E. Scharff. New York: Rowman & Littlefield.

Selesnick, S. A., Owen, G.S. 2012. "Quantum-Like Logics and Schizophrenia." *J Appl Logic*, 10: 115–26. https://doi.org/10.1016/j.jal.2011.12.001

Smith, D. L. 1999. *Freud's Philosophy of the Unconscious*. Dordrecht: Kluwer Academic Publishers. https://doi.org/10.1007/978-94-017-1611-6

Spillius, E. B., J. Milton, P. Garvey, C. Couve, and D. Steiner, eds. 2011. *The New Dictionary of Kleinian Thought*. London: Routledge. https://doi.org/10.4324/9780203815762

Steiner, R. 2003. "Introduction." In Steiner, R., ed., 2003. https://doi.org/10.4324/9780429484469-1

Steiner, R., ed. *Unconscious Phantasy*. 2003. London: Karnac. https://doi.org/10.4324/9780429484469

Svrakic, D. and Zorumski, C. 2021. "Neuroscience of Object Relations in Health and Disorder: A Proposal for an Integrative Model." *Front Psychol*, 12:583743, 1-25. https://doi.org/10.3389/fpsyg.2021.58374

Torok, M. 1994 [1968]. "The Illness of Mourning and the Fantasy of the Exquisite Corpse." In Abraham, N. and M. Torok, M. *The Shell and the Kernel: Renewals of Psychoanalysis*, edited by N. T. Rand, vol. 1, 107–24. Chicago: The University of Chicago Press.

VandenBos, G. R., ed. 2015. *APA Dictionary of Psychology*. Second Edition. Washington: American Psychological Association. https://doi.org/10.1037/14646-000

Vasylchenko, A. 2015. "Towards the Logic of Projective Identification." *J Appl Logic*, 13: 197–214. https://doi.org/10.1016/j.jal.2015.03.008

Von Domarus, E. 1944. "The Specific Laws of Logic in Schizophrenia." In *Language and Thought in Schizophrenia: Collected Papers*, edited by J. S. Kasanin. Berkeley: University of California Press.

Weiss, E. 1939. "The Psychic Presence." *Bull Menninger Clinic*, 3: 177–83.

Weiss, E. 1947. "Projection, Extrajection and Objectivation." *The Psychoanalytic Quarterly*, 16(3): 357–77. https://doi.org/10.1080/21674086.1947.11925686

Wollheim, R. 1984. *The Thread of Life*. Cambridge: Cambridge University Press.
Zizzi, P. and M. Pregnolato. 2012. "Quantum Logic of the Unconscious and Schizophrenia." *NeuroQuantology*, 10(3): 566–79. https://doi.org/10.14704/nq.2012.10.3.550

Organon F 32 (2) 2025: 150–167 ISSN 2585-7150 (online)
https://doi.org/10.31577/orgf.2025.32202

RESEARCH ARTICLE

You Don't Need Prime Matter: Welcome Rigid-Kooky Objects

Youssef Aguisoul*

Received: 22 January 2024 / Accepted: 4 January 2025

Abstract: The problem of destructive change comprises two interrelated questions: (1) Is there destructive change? (2) If there is, what underlies it? Classical hylomorphists argue that there is destructive change, understood as the change of primary substances, and that what underlies it is prime matter. Insofar as there is destructive change, I agree with classical hylomorphists. But there are reasons to doubt that prime matter is the underlying substratum, so I disagree with them with respect to (2). Alternatively, I propose a new version of classical hylomorphism, according to which what underlies destructive change, understood as the change of what I term "rigid-kooky" objects, is primary substance. My proposal has at least two perks. Compared to the classical account, it is relatively tolerant because it denies the historically contentious and ambiguous notion of prime matter. It is also economical because, unlike the classical account, which admits two constituency levels—one pertaining to primary substances and one to kooky objects—it admits only one constituency level, pertaining to kooky objects.

Keywords: Hylomorphism; prime matter; change; substantial change; accidental unities; substance.

* University of Lisbon
https://orcid.org/0000-0003-0377-8857
University of Lisbon, Cidade Universitária, Alameda da Universidade, 1649-004 Lisboa, Portugal
aguisoul@gmail.com

1. Introduction

Substantial or destructive change (let us treat these as synonyms for now) occurs when a primary substance (e.g., a cat) ceases to exist and transforms into another (e.g., a pile of ashes, assuming that piles are primary substances). But how is it possible for something to be generated out of something that ceases to exist? This problem has been addressed by traditional hylomorphists. What makes substantial change possible is a metaphysical postulate they call "prime matter."[1] When a cat, for example, ceases to exist, some prime matter remains—a bare stuff, roughly what Plato would call a "receptacle"[2]—ready to be actualized by the subsequent substantial form, in this case *being a pile of ashes*. What is prime matter? According to Aristotle, it is that which "is itself not a particular thing or quantity or anything else by which things are defined" (*Metaphysics* 1029a20); to Zeller, that "which is nothing but can become everything" (Zeller 1897, 247); to Oderberg, that which "is not an individual and it has no parts" (Oderberg 2022, 3). This postulate appears first in Aristotle's *Physics*, a book on the physically changing world. Ever since, it has been continuously defended and revised by Classical hylomorphists (e.g., Aquinas) and Neo-Classical ones (e.g., Stump 2003; Brower 2014; Oderberg 2007; 2022; Skrzypek 2024).[3] The aim of this paper is to motivate a new version

[1] Koslicki writes: "The traditional notion of prime matter is often traced back to Aristotle's views on substantial change..." (Koslicki 2021, 106).

[2] See *Timaeus* 48c–53c.

[3] The hylomorphism I focus on in this paper is Classical or neo-Classical hylomorphism. Therefore, I set aside what is known as Neo-Aristotelian or Mereological Hylomorphism (see e.g., Koslicki 2006; 2008; 2021; Fine 1999). Indeed, mereological hylomorphists typically reject prime matter. Writes Koslicki: "In defending a conception of matter that is able to capture the explanatory roles played by this notion in a hylomorphic account, hylomorphists have at least the following three options. First, the matter composing a hylomorphic compound can be construed along the lines of the traditional notion of "prime matter." The second option is to posit an ontological distinction between things or objects, on the one hand, and stuff, on the other hand, and to place the matter composing hylomorphic compounds in the second camp. According to the third approach, the matter composing a concrete particular object is analyzed not as prime matter or stuff, but rather as nothing more

of classical hylomorphism, which explains destructive change in the absence of prime matter. The general outline is as follows: In §2, I introduce basic hylomorphist tenets. In §3, I explain the problem of destructive change and argue that the postulation of prime matter only aggravates the issue. In §4, I propose an alternative solution deaf to prime matter, a solution that is relatively tolerant and economical. In §5, I conclude the paper.

2. Hylomorphism and Prime Matter

Hylomorphism resembles the contemporary doctrine of bare particularism.[4] Both substratum theories champion a kind of substratum, and both incorporate a realist constituent ontology.[5] According to bare particularism, concrete particular objects are constituted by bare particulars[6] and universals *tout court*. According to hylomorphism, they are constituted by prime matter

than the object's material parts which are themselves taken to be matter-form compounds and therefore to belong to the same ontological category, viz., that of concrete particular objects, as the whole they compose....Among the three options just cited, I have opted for the third, hylomorphic conception of matter in favor of the other two alternatives" (Koslicki 2021, 106).

[4] Proponents of bare particularism include: Alston (1954), Allaire (1963), Addis (1967); Bergmann (1967), Armstrong (1989), Moreland (1998), Sider (2006), Pickavance (2014), Wildman (2015), Paoletti (2023).

[5] Some believe that Aquinas's version of hylomorphism adopts a nominalist constituent ontology, while others believe differently. Here, I will simply assume that his account is realist. Witness Leftow: "Aquinas' theory of attributes is one of the most obscure, controversial parts of his thought. There is no agreement even on so basic a matter as where he falls in the standard scheme of classifying such theories: to Copleston, he is a resemblance-nominalist; to Armstrong, a 'concept nominalist'; to Edwards and Spade, 'almost as strong a realist as Duns Scotus'; to Gracia, Pannier, and Sullivan, neither realist nor nominalist; to Hamlyn, the Middle Ages' 'prime exponent of realism,' although his theory adds elements of nominalism and 'conceptualism'; to Wolterstorff, just inconsistent" (Leftow 2003, 1).

[6] Bare particulars, Bergmann writes, "neither are nor have natures. Any two of them are not intrinsically but only numerically different. That is their bareness. It is impossible for a bare particular to be 'in' more than one ordinary thing.... A bare particular is a mere individuator... It does nothing else" (Bergmann 1967, 24–25).

and *essential* universals or what hylomorphists call "substantial forms."[7] For example, while bare particularists say that a cat is constituted by a bare particular and accidental universals like *hairiness* as well as essential universals like *cathood*, hylomorphists say that it is constituted by prime matter and "only" *cathood*. Call this hylomorphic constitution of primary substances:

> *0-Constituency*. A primary substance is constituted by prime matter and a substantial form.

Unlike bare particularists, however, hylomorphists allow another constituency level. There is, as we just saw, the *bottom* constituency which includes prime matter and a substantial form, the combination of which gives rise to a primary substance. And there is the *upper* constituency which includes a primary substance and an accidental form, the combination of which gives rise to what hylomorphists call an "accidental unity" or a "kooky object":

> *1-Constituency*. A kooky object is constituted by a primary substance and an accidental form.[8]

Where is substantial change in all of this? According to hylomorphists, substantial change occurs, not in *1-Constituency*, but in *0-Constituency*. Does *1-Constituency* involve another type of change? Yes. It is what hylomorphists call "accidental change"—that type of change which involves no destruction, what Lowe (2006, 275) also calls "qualitative change," such as the change in which a cat, 'CAT,' goes from *being hairy* to *being groomed*. For better illustration, during accidental change, CAT loses *hairiness* over *being groomed*. CAT survives and the kooky object 'Hairy-CAT' perishes. CAT, behaving as a substratum, exits 'Hairy-CAT' and enters a new kooky object 'Groomed-CAT.' During substantial change, however, CAT perishes, as opposed to its prime matter which survives being a constituent of another primary substance, in this case a pile of ashes 'ASH.'

---

[7] The precise nature of substantial forms is currently a hot issue. For example, Rea (2011) takes substantial forms to be unifying powers; Marmodoro (2013) takes them to be space of possibilities. And Yates (2025) takes them to be geometric structures. Here, I naively read substantial forms as substantial or essential universals.

[8] More on kooky objects, refer to Matthew (1982), Lewis (1982, 29), and Cohen (2008).

Again, hylomorphists postulate prime matter to solve the problem of substantial change. Let this be:

> *0-Postulation*. Prime matter is postulated to solve the problem of substantial change.

Note that *0-Postulation* is related to *0-Constituency*. The latter depends on the former. If we do not "postulate" prime matter, then primary substances will not be "constituted" by prime matter.

In this paper, I argue that *0-Postulation* leads to an impasse, and that if no *0-Postulation*, then no *0-Constituency*. We are left with *1-Constituency*. I ask: is destructive change solvable via *1-Constituency*? I answer in the positive. The conclusion I make in this paper is that we do not need prime matter to solve destructive change. What is more, this conclusion turns out to be attractive. Since prime matter is, as many would say, historically contentious and ambiguous, and since explaining change in general (the non-destructive one, like CAT's changing from *being hairy* to *being groomed*, and the destructive one, like CAT's changing from *being a cat* to *being a pile of ashes*) via "two" levels of constituency is uneconomical, and since my proposal, according to which destructive change is solvable via *1-Constituency*, denies prime matter and explains both kinds of changes via only one level of constituency, it follows that my proposal is relatively tolerant and economical and, for that reason, should be taken seriously.

3. Prime Matter: Nowhere to Turn!

First, what is the problem of substantial change exactly? In my view, it comprises two interrelated sub-problems. The first can be formulated in terms of what I term *The Existence Question*: Is there substantial change? The second can be formulated in terms of what I term *The Persistence Question*: If there is substantial change, then what is the nature of its continuant substratum?[9] Hylomorphists answer *The Existence Question* in the

[9] Others, like Kronen et al, call this complex problem of substantial change "...the problem of the continuant....Traditionally the problem is set up so it looks as if it first must be settled whether there is substantial generation; that settled, by

positive. There is substantial change, say they, on the Aristotelian grounds that it is simply impossible for some primary substance to be generated out of absolute nothingness. Writes Oderberg:

> The quasi-creation and quasi-destruction that occur in the ordinary course of nature are familiar to both physicist and metaphysician: both the combustion of wood to particles of ash, and the decay of a muon into an electron, an electron antineutrino, and a muon neutrino, are instances of the same general process of substantial transformation. None of the entities involved reduces to absolute nothingness: energy/prime matter are simply transformed, literally, by the taking on of new forms and the 'reduction,' to use the Scholastic term, of the old forms back to potentiality" (Oderberg 2022, 8).

Proponents of anti-substantial change have several ways to defend themselves. For example, they might argue that primary substances can be generated out of absolute nothingness because miracles are possible. Arguing as such, they need not answer the *Persistence Question*. In their view, *nothing* persists after the destruction of primary substances: the generation of primary substances is grounded in miraculous states of affairs. By contrast, hylomorphists, who believe that substantial change is possible, must answer *The Persistence Question*. In their view, *something* persists after the destruction of primary substances: the generation of primary substances is grounded in the persisting thing at issue. What is that persisting thing? Prime matter.[10]

---

whatever argument, one then goes on to analyze the subject of the transformation" (Kronen et al 2000, 863–64).

[10] Oderberg's following quote summarizes the position of hylomorphists toward the two questions: "Of the various arguments for prime matter, the master argument is the one from substantial change: prime matter is a necessary metaphysical postulate because without it we cannot explain real substantial change [*The Existence Question*]...when I eat piece of celery the final products of digestion and metabolism are in no way numerically identical, either singly or collectively, to the piece of vegetable that entered my mouth. This latter example of substantial change is, metaphysically speaking, no different to the substantial transformation of particles in microphysics [*The Persistence Question*]" (Oderberg 2022, 3–4, my brackets).

One is tempted to suppose that prime matter, like any other substratum, explains identity through change. For example, just like CAT, as a substratum, explains the supposition that Hairy-CAT and Groomed-CAT are one and the same, one might suppose that the prime matter left after the destruction of CAT explains the supposition that ASH and CAT are one and the same. Hylomorphists tell us, though only by appealing to colloquial examples, not to take the latter supposition as true. Oderberg writes, "when I eat a piece of celery the final products of digestion and metabolism are in no way numerically identical, either singly or collectively, to the piece of vegetable that entered my mouth" (Oderberg 2022, 4). I doubt this, however, and shall come back to it shortly.

Now, is there substantial or destructive change? There is. So I agree with hylomorphists with respect to *The Existence Question*. Does prime matter explain it? I do not think so. So I disagree with hylomorphists with respect to *The Persistence Question*. I conclude that *0-Postulation*, according to which prime matter is postulated to solve substantial change, is dubious, and that, therefore, *0-Constituency* is dubious. To demonstrate my hearing, I assume first that prime matter explains substantial change. Then, from that assumption I derive an impasse.

Let us agree with hylomorphists that when a primary substance x of kind K is destroyed, there remains a continuant prime matter to render x's change to another primary substance y of kind K^* possible. Let us agree for example that there is a prime matter$_1$ which explains why CAT which is of kind *cathood* could change to this pile of ashes, ASH, which is of kind *ash-hood*. Let us illustrate this example via *0-Constituency*:

> At t_1, CAT is constituted by prime matter$_1$ and *cathood*.
>
> At t_2, ASH is constituted by prime matter$_1$ and *ash-hood*.

Now hylomorphists assume that ASH and CAT are numerically distinct (see Oderberg above). What makes them numerically distinct? An obvious part of what makes them so is that ASH is not a cat, nor a CAT is a pile of ashes. By comparison, since Hairy-CAT and Groomed-CAT are both of kind cat, they *might* be, given certain conditions,[11] numerically identical.

[11] If they also have the same substratum and accidental universals.

But since ASH and CAT fall into distinct "primary kinds," as Baker would put it,[12] this is sufficient to establish the conclusion that they are numerically distinct. So far, so good. But what of prime matter$_1$? If it is a substratum of CAT at t_1, and if it is the very same substratum of ASH at t_2, should not the conclusion also be that ASH and CAT are one in number? Compare. Hylomorphists would say that Hairy-CAT and Groomed-CAT are one in number since CAT figures as the underlying substratum of each at each time. Why a change of heart when it comes to prime matter? Hylomorphists, by the way, believe that a substratum of *any* change, substantial or accidental, accounts for the sameness of subject over time. Consider what Brower has to say on Aquinas's general account of change:

> ...worth noting about Aquinas's general account of change is that the notions of matter, form, and compound it employs are (at least in the first instance) purely functional in nature. To be matter, on this account, is just to be an entity playing a certain function or role—that of accounting for the sameness involved in change (namely, sameness of subject over time) (Brower 2011, 87).

But if matter, prime or otherwise qualified, accounts for the sameness involved in change, then CAT and ASH are ultimately one and the same subject. Perhaps hylomorphists will say that they are not on the grounds that whereas in accidental change CAT is a *particular*, in substantial change prime matter$_1$ is a *non-particular*.[13] And so while we can sensibly say that

[12] "For any x, we can ask: What most fundamentally is x? The answer will be what I call x's 'primary kind.' Everything that exists is of exactly one primary kind—e.g., a horse or a passport or a cabbage. An object's primary kind goes hand in hand with its persistence conditions. And since its primary-kind property determines what a thing most fundamentally is, a thing has its primary-kind property essentially: It could not exist without having its primary-kind property" (Baker 2002, 33–34).

[13] Koslicki writes, "prime matter is in and of itself not a particular thing nor does it belong to any of the other Aristotelian categories by which being is determined (viz., quality, quantity, etc.); it lacks an essence, since neither positive nor negative attributes belong to it in virtue of itself or its own nature … it is purely potential and cannot on its own exist actually; and, due to the fact that prime matter lacks an essence, it is therefore also in itself unknowable" (Koslicki 2021, 107; c.f. Oderberg 2022, 3).

Hairy-CAT and Groomed-CAT are ultimately one and the same subject or particular, we cannot sensibly say that CAT and ASH are ultimately one and the same non-particular. But if this is the reason, then it is not a good one. Numerical identity is not exclusive to particulars. Non-particular universals are numerically identical. Whatever the non-particularity of prime matter boils down to, then, nothing seems to prevent CAT and ASH from being regarded as one and the same. But if this is granted, and I do not see why it should not be, it would be incoherent to maintain that CAT and ASH are *numerically distinct* via the fact that they comprise distinct substantial forms, yet *numerically identical* via the fact that they comprise the same prime matter. The destroyed and the generated primary substance must be *either* numerically identical *or* distinct. Two available options before hylomorphists:

> (A) Hylomorphists could deny that prime matter explains identity through change, and admit the fact that two things having different substantial forms is sufficient to explain their numerical distinction. In which case, the destroyed and the generated primary substance come out numerically distinct *simpliciter.*
>
> (B) Hylomorphists could admit that prime matter explains identity through change, and deny the fact that two things having different substantial forms is sufficient to explain their numerical distinction. In which case, the destroyed and the generated primary substance come out numerically identical *simpliciter.*

Since hylomorphists themselves agree that substrata in general account for identity through change (see Brower above), they should want to opt for (B) rather than (A). If they do, however, they should accept that substantial forms like *cathood* and *ashhood* do not necessarily explain numerical distinction. Thus they should accept that ASH and CAT are ultimately one and the same subject despite the fact that the former is a pile of ashes and the latter is a cat. But this outcome seems to suggest that substantial forms behave much like accidental forms. For example, just like CAT remains one and the same when it loses *being hairy* for *being groomed*, it also remains one and the same when it loses *cathood* for *ashhood*. So, if hylomorphists opt for (B), they should accept that there is no genuine difference between

substantial and accidental forms, or, more precisely, they should accept a Quinean anti-essentialist stance and deny the privilege substantial forms have over accidental ones.[14] Hylomorphists, however, are essentialists *par excellence*. Brower, for example, writes:

> ...we can truly describe [lump] as a sphere at one time and a statue at another. Note, however, that the same cannot be said to hold of substantial changes, such as that involved in our human example. For a human to be generated from a zygote, there must be something—namely, some prime matter—that goes from *possessing zygotehood* to *possessing humanity*. Even so, such prime matter cannot itself be said to go from *being a zygote* to *being a human*. That is to say, we can't truly describe it as a zygote at one time and a human at another. And the reason has to do at least partly with the nature of the substantial forms or properties. Unlike *sphericity* or *statuehood*, *humanity* is not the sort of form or property that can characterize its possessor accidentally. On the contrary, it is a form or property that characterizes its possessor essentially: if something is *human* at any time it exists, it must be *human* at all (possible) times it exists. For the same reason, Aquinas thinks, it makes no sense to speak of something *coming-to-be* human—and likewise for *zygotehood* or any of the other forms involved in substantial change (Brower 2011, 90).

What is so bad about construing substantial forms as accidental? It certainly challenges our everyday intuitions. It invites us to doubt that these particular human beings are *essentially* human beings, and that these particular cats are *essentially* cats. It invites us to answer questions like 'what is Socrates?,' not in terms of 'Socrates is a human being,' but in terms of 'Socrates is potentially everything.'[15] Let us agree with hylomorphists that such invitations are below par. But if they are not ready to deny substantial forms, then they must be ready to opt for (A) instead of (B). Yet opting

[14] As W. V. Quine observes, essentialists have this "invidious attitude toward certain ways of uniquely specifying x … as somehow better revealing the essence of the object" (Quine 1953, 155).

[15] For a systematic discussion see Sullivan (2016).

for (A) involves the denial of substrata as entities that explain identity through change, something hylomorphists admit rather than deny.

Succinctly, if (A), no identity through change; if (B), no substantial forms; if neither (A) nor (B), then the transformed and the generated primary substance are numerically identical and distinct. Such an impasse is arrived at by means of the traditional hylomorphic answers to *The Existence Question* and *The Persistence Question*. Hylomorphists answer the former by saying that there is substantial change. And they answer the latter by postulating prime matter as the continuant of substantial change. If these answers lead to an impasse, the hylomorphist is called upon to put aside *0-Postulation*, according to which prime matter is postulated to solve substantial change, and thereby *0-Constituency*, according to which prime matter together with substantial forms constitute primary substances. The hylomorphist is called upon to entertain two possibilities: destructive change is explainable either via *0-Constituency* by way of postulating entities other than prime matter, or via *1-Constituency* alone. I think that the former option is not a desideratum. Instead of postulating further obscurities, one is better off considering whether *1-Constituency* can do the job. This option, as far as I know, has never been considered. I consider it here and motivate it against the classical hylomorphic account.

4. Kooky Change and Rigid-Kooky Objects

As an answer to *The Existence Question*, there is destructive change. But I do not understand the destructive transformation of, say, CAT to ASH as a transformation from one primary substance to another. Rather, I understand it as a destructive transformation from one kind of kooky object to another. At this point, I stop treating substantial and destructive change as synonyms. I take substantial change to be merely one type, among other types, of destructive change. It is a destructive change *with respect to primary substances*. There are other types of destructive changes. When day transforms into night, or when night transforms into day, we are probably talking about a destructive change *with respect to events*. Similarly, when an episode of happiness transforms into an episode of depression, we are probably talking about a destructive change *with respect to moods*. In

addition to these types of destructive change, there is a type *with respect to kooky objects*. And I call such destructive change "kooky change." I propose that when a cat transforms into a pile of ashes, what transforms and what generates are not primary substances but particular kinds of kooky objects. Thus, as an answer to *The Persistence Question*, what underlies a kooky change is a primary substance. Let me now unpack my proposal.

Now, what is the destructive change I call kooky change? Accidental changes occur in *1-Constituency*. At this level, a primary substance and an accidental form constitute a kooky object. For example, at t_1, CAT and the accidental form *hairiness* constitute Hairy-CAT. If at a time, t, a primary substance loses an accidental form, the former but not the latter survives. For example, suppose that at t_2, CAT loses *hairiness* in favor of *being groomed*. In that case, CAT survives and Hairy-CAT perishes; more specifically, CAT exits Hairy-CAT and enters a new kooky object, namely Groomed-CAT, and Hairy-CAT, but not CAT, undergoes a destructive change. I call that destructive change kooky change.

Here comes the heart of my proposal. Why not also add that primary substances can, as well, enter different kinds of kooky objects—objects we can term "rigid-kooky objects," such as 'Cat-CAT,' 'Ash-CAT,' 'Human-Socrates,' 'Frog-Socrates,' etc., as opposed to those like Hairy-CAT, Groomed-CAT, and 'Seated-Socrates,' which we can term "slack-kooky objects"? In this way, when we say that a cat is destroyed and a pile of ashes is generated, we mean to say that a rigid-kooky object, Cat-CAT, is destroyed and a pile of the rigid-kooky object, Ash-CAT, is generated. There is certainly ongoing destruction and generation! When CAT loses *cathood* over *ashhood*, it survives by exiting the destroyed rigid-kooky object, Cat-CAT, and entering the new rigid-kooky object, Ash-CAT. Before I demonstrate how my proposal avoids the impasse above and how attractive it is compared to the original hylomorphist version, let me make an important distinction between rigid-kooky and slack-kooky objects with respect to kooky change.

According to hylomorphists, kooky objects coincide with primary substances for a while.[16] Thus at t_1, CAT and Hairy-CAT coincide. Once we accept the possibility for 'two' things to coincide, nothing prevents us from

[16] See Cohen (2008, 4) and Pickavance (2014, 99).

also accepting the possibility for 'two or more' things to coincide. Hylomorphists accept that there are numerous coincidentals. So, if CAT at t_1 is hairy, grey, sleepy, and seated, then there are four slack-kooky objects coinciding with CAT at t_1, namely, Hairy-CAT, 'Grey-CAT,' 'Sleepy-CAT,' and 'Seated-CAT.' To this, I add that rigid-kooky objects can also coincide with primary substances and slack-kooky objects. Thus, at t_1, Cat-CAT coincides with CAT and the other four slack-kooky objects. Now, there is a crucial difference between a primary substance coinciding with a rigid-kooky object and it coinciding with a slack-kooky object. The difference is this: the former, but not the latter, involves a deep intimacy. That is, a primary substance like CAT coincides with a rigid-kooky object like Cat-CAT for a *long period of time*. Hence the term "rigid." By contrast, a primary substance like CAT coincides with a slack-kooky object like Hairy-CAT or Seated-CAT for *a short period of time*. Hence the term "slack." We can call these coincidences respectively "rigid-coincidence" and "slack-coincidence." Such coincidences explain why primary substances retain a particular form for a good while, while losing others in a flash.

My proposal avoids the impasse in Section 3. Hylomorphists, who explain destructive change—understood as a destructive transformation of primary substances—reach the trilemma that the destroyed and the generated primary substance are numerically identical and distinct; if only identical, no substantial forms; and if only distinct, prime matter is unable to explain identity through change. My proposal, which explains destructive change—understood as a destructive transformation of kooky objects—does not reach such a roadblock. For, *1-Constituency*, the level at which kooky changes occur, consists, not of prime matter and substantial form, but of primary substance and accidental form. Here, the primary substance, being the substratum, accounts for identity through change, and so the destroyed and the generated rigid-kooky object are ultimately one and the same subject, e.g., Ash-CAT and Cat-CAT are one and the same ultimate subject, CAT. Moreover, the rigid-kooky objects are not numerically distinct via the supposition that each falls under a distinct primary kind or substantial form. For, again, such objects are not constituted by substantial forms. Ash-CAT is constituted by the accidental form *ashhood* and Cat-CAT by *cathood*.

One pressing objection may be that my proposal wrongly reduces substantial forms to accidental forms. For example, it treats *cathood* as accidental, yet *cathood* is clearly substantial. This objection would be correct absent some clarification on my part. Once my proposal is clarified, the objection will lose some of its force.

Recall that the original hylomorphist account involves two constituencies, each containing a particular type of form. In *0-Constituency*, there are substantial forms. In *1-Constituency*, there are accidental forms. My proposal denies *0-Constituency*. Does that mean I deny substantial forms? No. Then what happens to them? I locate them in *1-Constituency*. Does that mean that kooky objects are now constituted by primary substances, substantial forms, and accidental forms? No, for it depends on which kooky object is under discussion. If we are talking about slack-kooky objects, then each of these is constituted by a primary substance and an accidental form (e.g., *hairiness*). If we are talking about rigid-kooky objects, then each of these is constituted by a primary substance and a substantial form (e.g., *cathood*).

Still, one might correctly observe that my account undermines the distinction between substantial and accidental forms. For whether we are discussing slack-kooky or rigid-kooky objects, both are destroyed when their associated form is destroyed. Hairy-CAT perishes once *hairiness* perishes, and Cat-CAT perishes once *cathood* perishes. My account controversially takes *hairiness* to be as substantial as *cathood*, and *cathood* as accidental as *hairiness*.

Accordingly, I choose to construe the nature of forms as *context-sensitive*, in the sense that their substantiality or accidentality entirely depends on the subject under discussion. If primary substances are the context, then forms are accidental. If kooky objects are the context, then forms are essential. So instead of *1-Constituency*, I propose:

> *1-Constituency**. A kooky object is constituted by a primary substance and a *context-sensitive* form.

For example, Hairy-CAT is constituted by CAT and context-sensitive *hairiness*, such that the latter is accidental with respect to CAT, but substantial with respect to Hairy-CAT. Similarly, Cat-CAT is constituted by CAT and context-sensitive *cathood*, such that the latter is accidental with respect to

CAT, but substantial with respect to Cat-CAT. But if *1-Constituency**, then I do not undermine the distinction between substantial and accidental forms, for such forms are *not* context-free.[17]

My new version of classical hylomorphism has its perks. For one, traditional hylomorphism takes the nature of forms to be context-free. Thus it considers *cathood* to be substantial *simpliciter* and *hairiness* accidental *simpliciter*. However, this construal faces difficulties when considering how subjects have their forms. In this traditional view, one is allowed to say that a primary substance has its substantial form substantially (e.g., CAT has its substantial form, *cathood*, substantially) and its accidental form accidentally (e.g., CAT has its accidental form, *hairiness*, accidentally). This double-having of forms can be problematic,[18] and it is, above all, convoluted compared to an account that admits only the single-having of forms. My proposal, which takes the nature of forms to be context-sensitive, eliminates the double-having. A primary substance has a form accidentally, while a kooky object has a form substantially.

For two, my version explains kooky change—whether rigid-kooky change or slack-kooky change—by considering only one level of constituency, namely, *1-Constituency*. This is economical. Traditional hylomorphism, in contrast, is uneconomical in this respect.

Finally, my version denies prime matter, an exceedingly contentious postulate. Thus, my version is exceedingly tolerant.

5. Conclusion

I admit that entities like cats and atoms undergo destructive transformations, but I deny that these entities are primary substances. Therefore,

[17] The idea of having a form accidentally or essentially bears resemblance to the distinction proposed by bare particularists, namely, between internal exemplification (having a property essentially) and external exemplification (having a property accidentally). See Alston (1954, 257); Connell (1988, 90); Moreland (1998, 257); Pickavance (2014). Using that distinction, Cat-CAT would internally exemplify cathood, while CAT would externally exemplify it.

[18] See e.g., Bailey (2012).

I deny what hylomorphists call substantial change. I proposed that we construe such entities as kinds of kooky objects—objects I termed rigid-kooky objects. I suggested that we substitute rigid-kooky change for substantial change. This proposal has its share of attractions. Compared to mainstream hylomorphism, my version is economical (it explains destructive change using only *1-Constituency*) and exceptionally tolerant (it denies prime matter, which has historically been contentious and ambiguous).

To be sure, my version is not without obscurities. For one, it takes primary substances to be bare particulars—particulars devoid of their own forms. Recall that CAT is not itself *being a cat* or *being hairy*; such forms characterize rigid-kooky objects and slack-kooky objects, respectively. But is it plausible to construe primary substances as bare particulars? Such an identification has been proposed recently. Connolly (2015), for example, takes primary substances to be identical to bare particulars. The difference between his view and mine, however, is that, whereas he denies that bare particulars/primary substances are constituents of anything, I admit that they are constituents of slack-kooky and rigid-kooky objects. Are both views implausible, given their identification of primary substances with bare particulars? And if such identification is acceptable, then which view is more palatable—Connolly's or mine? These issues are for another day. Today, I bid farewell to the long-standing prime matter and the long-standing 0-Constituency.[19]

References

Addis, Laird. 1967. "Particulars and Acquaintance." *Philosophy of Science*, 3(3): 251–59. https://doi.org/10.1086/288156

Allaire, Edwin B. 1963. "Bare Particulars." *Philosophical Studies*, 14, (1/2): 1–8. https://doi.org/10.1007/BF00396663

Alston, William P. 1954. "Particulars--Bare and Qualified." *Philosophy and Phenomenological Research*, 15(2): 253–58. https://doi.org/10.2307/2103580

Aquinas. 1965. *On Being and Essence*, translated by Joseph Bobik. Notr Dame, Indiana: University of Notre Dame Press. https://doi.org/10.2307/j.ctvpj7dxh

[19] The central ideas of this paper would not have emerged sans the guidance of David Yates and Bruno Jacinto, to whom I am profoundly grateful.

Armstrong, David. 1989. *Universals: An Opinionated Introduction*. Boulder, Colorado: Westview Pres. https://doi.org/10.2307/2186070

Bailey, Andrew M. 2012. "No Bare Particulars." *Philosophical Studies*, 158(1): 31–41. https://doi.org/10.1007/s11098-010-9665-2

Baker, Lynne Rudder. 2002. "The Ontological Status of Persons." *Philosophy and Phenomenological Research*, 65(2): 370–88. https://doi.org/10.1111/j.1933-1592.2002.tb00207.x

Bergmann, Gustav. 1967. *Realism: A Critique of Brentano and Meinong*. Madison, Wisconsin: University of Wisconsin Press.

Brower, Jeffrey E. 2011. "Matter, Form, and Individuation." In *The Oxford Handbook of Aquinas*, edited by Davies Brian and Stump Elenor, 85–103. Oxford: Oxford University Press.

Brower, Jeffrey E. 2014. *Aquinas's Ontology of the Material World: Change, Hylomorphism, and Material Objects*. Oxford: Oxford University Press.

Brower, Jeffrey E. 2017. "Aquinas on the Individuation of Substances." *Oxford Studies in Medieval Philosophy*, 5(1): 122–50. https://doi.org/10.1093/oso/9780198806035.003.0004

Cohen, Mark S. 2008. "Kooky Objects Revisited: Aristotle's Ontology." *Metaphilosophy*, 39(1): 3–19. https://doi.org/10.1111/j.1467-9973.2008.00521.x

Connolly, Niall. 2015. "Yes: Bare Particulars!" *Philosophical Studies*, 172(5): 1355–70. https://doi.org/10.1007/s11098-014-0353-5

Connell, Richard. 1988. *Substance and Modern Science*. Notre Dame, Indiana: University of Notre Dame Press.

Koslicki, Kathrin. 2006. "Aristotle's Mereology and the Status of Form." *The Journal of Philosophy*, 103(12): 715–36. https://doi.org/10.5840/jphil2006103127

Koslicki, Kathrin. 2008. *The Structure of Objects*. Oxford: Oxford University Press.

Koslicki, Kathrin. 2021. "Form, Matter, Substance." *Chroniques Universitaires*, 2020, 99–119.

Kronen, John D., Sandra Menssen, and Thomas D. Sullivan. 2000. "The Problem of the Continuant: Aquinas and Suarez on Prime Matter and Substantial Generation." *The Review of Metaphysics*, 53(4): 863–85.

Leftow, Brian. 2003. "Aquinas on Attributes." *Medieval Philosophy and Theology*, 11(1): 1–41. https://doi.org/10.1017/s105706080300001x

Lewis, Frank A. 1982. "Accidental Sameness in Aristotle." *Philosophical Studies*, 42(1): 1–36.

Lowe, Edward J. 2006. "How Real is Substantial Change?" *The Monist*, 89(3): 275–93.

Marmodoro, Anna. 2013. "Hylomorphism without Reconditioning," *Philosophical Inquiry*, 37(1–2): 5–22. https://doi.org/10.5840/philinquiry2013371/28

Matthews, Gareth B. 1982. "Accidental Unities." In *Language and Logos*, edited by Schofield Malcom and Nussbaum Martha, 223–40. Cambridge: Cambridge University Press.

Moreland, James P. 1998. "Theories of Individuation: A Reconsideration of Bare Particulars." *Pacific Philosophical Quarterly*, 79(3): 251–63. https://doi.org/10.1111/1468-0114.00061

Oderberg, David. 2007. *Real Essentialism*. New York: Routledge.

Oderberg. David. 2022. "Is Prime Matter Energy?" *Australasian Journal of Philosophy*, 101(3): 1–17. https://doi.org/10.1080/00048402.2021.2010222

Paoletti, Michele P. 2023. "Bare Particulars, Mode, and the Varieties of Dependence." *Erkenntnis*, 88(4): 1593–1620. https://doi.org/10.1007/s10670-021-00417-6

Pickavance, Timothy. 2014. "Bare Particulars and Exemplification." *American Philosophical Quarterly*, 51(2): 95–108.

Rea, Michael C. 2011. "Hylomorphism Reconditioned." *Philosophical Perspectives*, 25(1): 341–58. https://doi.org/10.1111/j.1520-8583.2011.00219.x

Quine, Willard Van Orman. 1953. "Reference and Modality." In *From a Logical Point of View*, 139–59. Cambridge, Massachusetts: Harvard University Press.

Sider, Ted. 2006. "Bare Particulars." *Philosophical Perspectives*, 20(1), 387–97. https://doi.org/10.1111/j.1520-8583.2006.00112.x

Skrzypek, Jeremy W. 2024. "Thomas Aquinas on Concrete Particulars." *American Catholic Philosophical Quarterly*, 98(1): 49–72. https://doi.org/10.5840/acpq2024515288

Stump, Eleonore. 2003. *Aquinas*. New York: Routledge.

Sullivan, M. 2016. "Are There Essential Properties? No." In *Current Controversies in Metaphysics*, edited by Barnes Elizabeth. New York: Routledge.

Wildman, Nathan. 2015. "Load Bare-ing Particulars." *Philosophical Studies*, 172(6): 1419–34. https://doi.org/10.1007/s11098-014-0356-2

Yates, David. 2025. "Hylomorphism, or Something Near Enough." In *Rethinking Emergence*, edited by David Yates and Amanda Bryant. Oxford: Oxford University Press.

Zeller, E. 1897. *Aristotle and the Earlier Peripatetics*. New York: Russell & Russell.

Organon F 32 (2) 2025: 168–192 ISSN 2585-7150 (online)
https://doi.org/10.31577/orgf.2025.32203

RESEARCH ARTICLE

Hypothesis Testing and Knowledge of the Human Past

David Černín*

Received: 22 January 2024 / Revised: 24 March 2025 / Accepted: 31 March 2025

Abstract: Historians and natural scientists are adept at inferring knowledge of the past from present traces and evidence. The repository of available methods has been rapidly expanding, and historians of the human past have learned that using techniques developed in other fields that study the natural past might prove beneficial to their endeavours in some cases. The network of inferences involved in historical discourse is vast and diverse. Influential philosophers of history and historical sciences like Jouni-Matti Kuukkanen and Aviezer Tucker have argued that it is necessary to expand the scope of the philosophy of history and to take a deeper look at underlying reasoning and inferential structure of history. This paper answers this call by analysing a group of inferences in human historiography that could be described in terms of hypothesis and hypothesis testing. Hypothesis testing has received some attention in the philosophy of historical sciences, but it is mostly underexplored in the philosophy of history in the context of human history. The paper examines four case studies and analyses their inferential structures by using concepts from the philosophy of historical sciences, such as trace-based and analogous reasoning, type/token distinction, etc. It will be shown that hypothesis testing helps generate knowledge about the past, but

* Ostravská univerzita
https://orcid.org/0000-0001-8711-3929
Katedra filozofie, Filozofická fakulta, Ostravská univerzita, Reální 5, 701 03 Ostrava, Česká republika
David.Cernin@osu.cz

to fully appreciate it and to differentiate its types, philosophers of history must engage with the infrastructure of historical research.

Keywords: Hypothesis testing; evidence; experimental archaeology; intellectual history; philosophy of historiography.

1. Outset

Can we arrive at a knowledge of the past via testing hypotheses, and how much can we learn in this manner? In the last few decades, this question has been mostly explored in the realm of historical natural sciences. For example, Derek Turner (2007; 2013) and Carol Cleland (2001, 2013) have engaged in numerous debates regarding the role of prediction and hypothesis testing in geological research. Whereas Cleland argues that *historical sciences* are interested in inferences about token events from the present evidence and they are largely distinct from *experimental sciences*, which operate via using hypothesis and prediction, Turner understands experimental methods and hypothesis testing as a part of a basic toolbox of historical sciences, which can make use of regularities between *type* events (see also Jeffares 2008, Tucker 2011, or Currie 2018).

In contrast, the process of hypothesis testing in human history, historiography, intellectual history, and archaeology remains comparatively unexplored,[1] even though an inquiry into its logic might prove fruitful for our philosophical understanding of historical discourse and its inferential structures. In 1968, Rolf Gruner published a paper explicitly dealing with hypothesis testing in history:

> In short, saying that historians test hypotheses when they establish facts is, if not outright false, at least very misleading. It cannot be more than an analogy, but not a good one. A better analogy would be to speak of a pattern of interlocking pieces which

[1] Hypothesis testing receives attention especially in relation to archaeology, see, e.g., "Central Place Theory and the Reciprocity between Theory and Evidence" by Peter Kosso and Cynthia Kosso (1995) or book *Evidential Reasoning in Archaeology* by Robert Chapman and Alison Wylie (2016).

> are fitted together similar to the fitting together of a jigsaw puzzle and where the pieces consist of facts of evidence and known historical facts (Gruner 1968, 128).

Much more recently, yet another point of view was offered by the rise of the non-representationalist or postnarrativist philosophy of history. While previous discussions revolved around historical explanations and narratives, the focus shifted towards historical discourse as an argumentative practice. In the words of Jouni-Matti Kuukkanen, the main question is: "How did a historian X arrive at his view of the past given his argumentative context, the sources and the texts available to him?" (Kuukkanen 2017, 118). In 2021, he pursued this project in a text called "Historiographical Knowledge as Claiming Correctly." Here, he provides a case study of the book *Revolt at Factories* by Seppo Aalto and explores how selected statements from this book are warranted. The background question of this text is whether history is an empirical discipline and what constitutes evidence. Kuukkanen calls for a deeper study of the inferential structures, how historians provide justification for their statements in practice, and how historians decide between legitimate inferential moves and illegitimate ones.

Kuukkanen differentiates between several types of inferences or linguistic acts performed by historians, which he could identify in the book he examined for his case study: "inference from archival material, inference from literature, inference from shared beliefs (historiographical and moral), textual inference, textual coherence, the authority of the historian" (Kuukkanen 2021, 63). Given that this list is founded on one particular case study and one particular piece of historiography, it can not be considered exhaustive, and Kuukanen does not claim that it is. I do believe that the study of inferential practices that historians employ in their research is a fruitful endeavour that should be pursued in addition to other lines of inquiry in the philosophy of history. However, in this paper, I will claim that we need to expand the scope and depth of our dive into the intricacies of historical inference.

While close reading of historiographic texts may reveal a lot about the logic of historical discourse, it begs several questions. First, if philosophers of history claim to analyse historical discourse, they should be clear about the scope of such enterprise. Should we focus predominantly on extensive

synthesising pieces with clear narrative structure, explicitly worded explanations, and on the reception of these accounts? Or are we to be more inclusive and analyse smaller building blocks of this discourse, like individual papers that focus on establishing the authenticity or relevance of a particular piece of evidential record and focus on minute details in microhistorical studies? What is the role of complex ancillary disciplines like archaeology, aDNA analysis, or experimental ethnographic studies, and how do they factor in the complex network of historiographic discourse and its inferential structures? I would not expect serious objections against a claim that these intellectual activities constitute historical discourse and that we should include written outcomes of these activities in our philosophical reflection of the field.

The second question urges us to go even deeper: Are there inferential activities that are not readily identifiable in every written output of historical discourse regarding a certain topic? Leon J. Goldstein coined a distinction between the infrastructure and the superstructure of history. He defined the superstructure as "that part of the historical enterprise which is visible to nonhistorian consumers of what historians produce" (Goldstein 1976, 141). Thus, the superstructure encompasses literary products (and possibly other types of media) that are the most visible part of historical discourse. Goldstein considers the infrastructure of history to be crucial for philosophical reflection since

> it involves treatment of evidence and thinking about evidence and is preoccupied with the determination of what conception of the historical past makes best sense given the character of the evidence in hand (Goldstein 1976, 141).

Even more importantly, Goldstein specifies (1986, 87) that the infrastructure of history concerns activities that are often missing from the final accounts, or they appear only in footnotes, if at all.

Kuukkanen is suspicious of Goldstein's attempt to draw a thick line between the superstructure and infrastructure:

> the distinction between superstructure and infrastructure is not solid, because presentation is a part of the justification of a historiographical work and therefore must be a subject of historiographic epistemology (Kuukkanen 2015, 7).

It could be argued that the superstructure of history exerts a notable power on historians' practice at the infrastructure level. A decision to write a historical account presupposes some idea of the past, knowledge of the discourse and its rules, accepted frameworks, and contemporary terminology. Thus, the preliminary idea about the subject matter may influence the search for evidence and other intellectual activities that Goldstein associates with the infrastructure.

Nevertheless, Goldstein's call for greater attention towards the inferential structures hidden below the surface of textual results of historian's work remains. Interestingly, Goldstein described these intellectual activities in terms of hypothesis testing (1986, 88). The paper will thus focus on these types of intellectual operations in the process of historical discourse and will try to describe their logic.

By a hypothesis in the context of history and archaeology, I mean a statement about the past that receives some degree of support from other theories in the field, but the support is not sufficient to warrant its uncontested acceptance in historical discourse. In experimental sciences, it is easy to accept Peter Kosso's description: "A hypothesis is a theory that has little testing and is consequently located near the speculation-end of the spectrum" (Kosso 2011, 8). A historical hypothesis is thus viewed as speculation unless additional research activities can somehow decrease the uncertainty, i.e., to support or infirm a hypothesis. "Statements about unobservable things can be tested by their observable implications" (Kosso 2011, 13). Kosso is clear that these implications may include traces of the (unobservable) past, like fossils (2011, 14). In these cases, testing involves searching for additional traces that may serve as evidence, thus increasing consilience. However, we can identify other instances of research activities that may either increase the informational value of available traces and evidence, or produce entirely new evidential bases for the claims about the past, like experimental archaeology or ethnoarchaeology.

This paper will focus on differences between various instances of testing historical hypotheses. The main focus will be on the testing process itself and the bearing of different results on warranted claims about the past, while the inferential activities involved in formulating the hypothesis will be mentioned only where necessary.

2. Evidence-Seeking Hypothesis Testing

Broad historical narratives do not cover the process of historical research in detail. Should we use Goldstein's language, the infrastructure is mostly hidden, except for occasional references to archival material. Authors presuppose (and they are mostly correct) that a general reader is not interested in day-to-day groundwork. In some cases, the context of the discovery of a certain relict or text might be amusing or revolutionary enough to receive a mention in the text or in the footnote. However, if we really are to argue for "practice revolt" (Kuukkanen 2021, 64), then philosophers should search deeper for underlying inferential structures.

The emblematic workspace of a historian is an archive. Historians go there with a certain research question in mind (especially when they need to justify a trip and associated costs according to the guidelines of a grant agency). Be it a study focusing on the life of a particular person or a loftier goal of shedding more light on some pivotal historical event. Archives are a product of peculiar social practice, as noted by Robin G. Collingwood during the dawn of the philosophy of history:

> ...the past leaves relics of itself, even when these relics are not used by any one as materials for its history; and these relics are of many kinds, and include the relics of historical thought itself, that is, chronicles. We preserve these relics, hoping that in the future they may become what now they are not, namely historical evidence" (Collingwood 1994, 203).

Goldstein, himself being loosely influenced by Collingwood, holds that evidence must always be relative to some previously conceived theory or hypothesis (Goldstein 1962, 180), and Kuukkanen's view of evidence does not seem to stray away, since evidence "can then in general be understood as anything that makes something reasonable to be believed" (Kuukkanen 2021, 64). Yet none of this seems to be at odds with a claim that historians are adept at recognising traces or relicts of the past even before they are identified as evidence for some specific claims or theories about the past, and they participate in practice to preserve, store, and catalogue the traces of the past. In general, it could be argued that traces of the past have some informational value about their origins (Tucker 2025, 4). As historians develop their

methods, they become better at decoding the information contained within the traces, or they learn to recognise additional *types* of traces. These traces may or may not be used as evidence for novel claims about the past. The utilisation of traces and information they contain as evidence is conditioned by research questions that may arise during historical research.

Let us illustrate this process: In the book *Truth and History*, philosopher Murray G. Murphey provides an extensive example of research he was participating in (Murphey 2009, 40–43). The research itself consisted of reconstructing the life of a particular Wyoming physician. Murphey describes this research in painstaking detail and goes over every aspect that baffled historians: sudden changes in the name (Hart/O'Hart) and the date of birth (1890/1891) the historical agent himself reported in archival documents. Murphey focuses on the process of how every discovery provided both some new hypothesis and a clue as to where to look next for another trace – the subsequent potential evidence. It is possible to quote a comparatively brief passage that exhibits all these steps: formulating a likely hypothesis based on available data, identifying ways of testing the hypothesis, testing the hypothesis via further inquiries, and evaluating the results with regard to the original hypothesis.

> But we needed to be sure James O'Hart was the father of Hart/O'Hart rather than Patrick. Nebraska did not institute birth certificates until 1904, but if the O'Harts were Catholic, then there should have been a baptismal record. From the Catholic Register for 1890 we found that in 1890 the nearest Catholic Church to Murray was in Plattsmouth—the county seat, and there at the Church of the Holy Spirit we found a baptismal certificate certifying that James Benedict O'Hart had been baptized in June 1890 with parents James O'Hart and Mary Ann Quinn O'Hart and witnesses Patrick O'Hart and Catherine Quinn. So we were now sure that James Oakes Hart was really James Benedict O'Hart, son of James O'Hart, and it was a fair guess that Catherine Quinn was his mother's sister and probably became Patrick O'Hart's wife. We were able to confirm that from a newspaper notice of their marriage in December 1890 (Murphey 2009, 42).

To summarise, Murphey states:

> At every point, the historian asks himself, if the situation was as I think it was, what should I be able to find, and where should I be able to find it? It should be obvious that the process consisted of a series of predictions and inferences as to what and where the data would be that were then followed up and confirmed or infirmed (Murphey 2009, 43).

In the context of the chapter, Murphey is trying to prove that historians are able to make predictions, i.e., they predict what data they are trying to find and where they will find them; similarly to scientists who are essentially predicting what they will observe under certain conditions (Murphey 2009, 45). However, more interesting is Murphey's description of inferential processes guiding historians from one archive to another, their ability to imagine potential evidence without possessing it and inferring the most probable location of said evidence. This is made possible because historical discourse (as a collection of guidelines and commonly accepted practices) provides historians with some generalised idea of historical evidence.[2] Furthermore, we have developed a custom of collecting and archiving texts and artefacts that might serve as evidence in the future. Equipped with historical training and awareness of our archival practices regarding documents and artefacts, a historian can propose a testable hypothesis that if x was a case, then there might be evidence y located at $z_1 \ldots z_x$.

This (1) *evidence-seeking hypothesis testing* is common throughout historical practice. In some cases, it might direct historians towards previously unconnected parts of contemporary professional historical discourse (e.g., when comparing two well-documented processes that were once considered unrelated). In other cases, it might require archival research (e.g., tracing intellectual influence among medieval scholars requires browsing through unedited original manuscripts and assessing what sources they might have

[2] The very fact that historians possess some generalised idea of what can be historical evidence is a necessary condition for an insidious practice of creating forgeries and fake artefacts (like the Kensington Runestone) in order to pursue personal or ideological goals. To counter this practice, historians are taught to strictly scrutinise any potential evidence.

had at their disposal, whether they possessed full texts or mere fragments, etc.). (1) *Evidence-seeking hypothesis testing* is an essential part of historians' daily work.

As such, this type of hypothesis is not often explicitly stated in the finished product (e.g., historical narrative), which usually presents a tidy account of past events with a complete list of sources and historical records. Goldstein observed that evidence for various historical claims often appears in footnotes; however, he also acknowledged that even footnotes do not present the entire argumentative background behind the finished account (Goldstein 1986, 87). In the same vein, Kuukkanen states:

> In order to study the inferential structure, it is necessary to study the nitty-gritty and follow inferential chains and networks wherever they lead in order to see from where the reasonableness of historians' claiming stem (Kuukkanen 2021, 64).

To pursue these goals, it is imperative to look beyond finished historical accounts and to analyse historians' legwork and their ability to make hypotheses and seek evidence.

We should examine another aspect of the (1) *evidence-seeking hypothesis testing,* which has not been covered extensively by previously mentioned philosophers. When a historian makes a reasonable hypothesis based on available evidence about past events (a historical agent changing the name, intellectual influence between two medieval scholars, social tensions preceding border conflicts), she should be able to postulate possible evidence which would make her hypothesis more probable and its location in order to maintain epistemic diligence. Her subsequent journey to archives or libraries serves as a way to test the hypothesis. Either her research among historical documents will yield results, or she will not find predicted evidence. In a favourable scenario (e.g., a discovery of a letter explaining historical agent's intentions in changing the name; a finding of a copy of a particular historical document authored by an earlier scholar included in a broader unedited manuscript written by a later medieval scholar, which may serve as evidence for intellectual influence) the hypothesis receives some backing.

In a negative case (evidence was not found),[3] the hypothesis was neither confirmed nor thoroughly falsified. The hypothesis was "infirmed", as Murphey puts it. The hypothesis about the past might still be plausible or even true, but the evidence was not found and observed, which can mean either that it was not preserved or is located elsewhere. One aspect that Murphey does not discuss in detail is the degree of changes in the probability of a hypothesis given the evidence (or its absence). Suppose we follow Kuukkanen's call for a deeper study of inferential processes. In that case, we should strive for further examination and learn how exactly different outcomes of this process move a needle regarding the probability of the hypothesis.

Here, significant progress was achieved by philosophers who employ the Bayesian probabilistic framework (Sober 2009; McGrew 2014; Tucker 2025) and discussed how evidence or its absence impacts the probability of a hypothesis. Going back to Murphey's example, we understand that if we find a baptismal record in the archive, we were able to pinpoint, with the help of the original hypothesis, that it is highly improbable that the record (evidence) would be observed, whilst the hypothesis would be false.[4] The absence of evidence or – more accurately – our failure to observe the evidence also moves a needle, but its impact is much weaker (Sober 2009, 88–89). McGrew investigated the cases of absence of evidence in relation to historiography, and he noted that if we hypothetically suppose some event to be true, we need to approximate additional probabilities, like how probable it is that somebody would have noticed the event, how probable it is that the event would have been recorded, and how probable is the survival of the given record for a historian to observe (McGrew 2014, 221–222). In the case of the baptismal record as a well-documented established practice, the absence of such a record if the hypothesis is true would seem improbable; thus,

[3] In some cases, it is also possible to find data (texts, documents, photos) that directly contradict the original hypothesis. In such cases, the hypothesis should be considered severly infirmed and a different claim about the past should be considered strengthen by new evidence. Such a result is more akin to the abovementioned positive case.

[4] Barring other competing explanations that might be investigated independently, e.g., the baptismal record is a forgery.

the absence of the record will notably decrease the probability of the hypothesis as a whole. However, there are other cases where the absence of a record might not be as decisive. McGrew (2014, 224–225) investigates an interesting case of an absent historiographic record of a major conflagration in Bergen in the early 13th century that is otherwise sufficiently evidenced via the archaeological record. The apparent failure of contemporary chroniclers who did not record the fire is explained via a reference to a theory regarding the record-keeping practices of medieval chroniclers, who often recorded similar disasters selectively and in the context of broader narrative demands (Dunlop and Sigurdsson 1995, 87–89). In other words, the absence of written evidence about the fire is not surprising enough to outweigh the archaeological record and infirm the hypothesis concerning a major conflagration in Bergen because contemporary chroniclers were selective about the fires they recorded. In the more recent past, people may decide not to keep some types of records or to hide them, e.g., because of a fear of persecution by a totalitarian regime (Tucker 2025, 55). Thus, to the result of (1) *evidence-seeking hypothesis testing*, a vast array of theories is necessary, including the theories about the practices of records-keeping in the past.

To summarise, Goldstein urges us to take a look at the infrastructure of historical discourse that is often missing from the final narrative outcomes. With Murphey, he views some of these intellectual practices as a form of hypothesis testing. Following their proposal, we can see that by assessing the impact of evidence (or its absence) on a hypothesis, we uncover another rich network of inferential practices below the simple statements about the past and (in some cases) these practices entail trace-based reasoning and informational epistemology (e.g., Dunlop and Sigurdson 1995 established a theory regarding practices of medieval chroniclers via a sufficiently robust comparative study of numerous traces). Thus, theories concerning transmission and loss of information are crucial in any project that strives to take historical discourse as a serious and rational endeavour practised by a group of experts that is open to warranted revisions. "The general epistemological point is that any claim that can justify (test and verify) what we claim to know about the past must itself be justified" (Kosso and Kosso 1995, 593).

3. Experimental Testing of a Historical Hypothesis

Let us now turn to another discipline, which is, on the one hand, interested in the human past, but, on the other hand, it explicitly frames its methodology in hypothesis testing. There is perhaps no better candidate than *experimental archaeology*. According to Alan K. Outram and his "Introduction to Experimental Archaeology" (Outram 2008, 2), this line of research should be seen as a natural follow-up from archaeological laboratory experiments that inquire into the properties of materials. According to the guidelines proposed by prominent experimental archaeologist Peter Reynolds, a hypothesis is formed on the basis of archaeological data alongside available documentary sources (Reynolds 1999, 127). For example, we may form hypotheses regarding the manufacturing process by analysing ancient pottery and its fragments coupled with historical texts. Such hypotheses can be tested via physical experiments and their replications. In this process, two sets of data are distinguished—original data (i.e., archaeological record and its properties) and data produced by the experiment. Reynolds then states:

> If there is agreement between the sets of data, the hypothesis can be tentatively accepted as valid but with the caveat that several different hypotheses raised on the same data might also be validated. If there is no agreement, the hypothesis is not merely invalidated but actually proved to be wrong. The value of this methodology lies especially in the seemingly worst case situation (Reynolds 1999, 127).

This crudely Popperian statement is central to the guidelines of experimental archaeology; however, we may consider it too blunt. Failure to confirm the hypothesis through a well-designed experiment does not lead to invalidation in a straightforward manner (Cleland 2001, 988). It would be more accurate to say that the hypothesis did not receive meaningful support from the experiment. The hypothesis was infirmed and is likely to be rejected by professionals.

An experiment successful at replicating the identified properties of archaeological record shows that this particular approach to manufacturing *might* have been used (leaving space for alternative hypotheses). In contrast,

the failed experiment may substantially infirm the hypothesis, provided that the experiment was designed correctly.

Reynolds himself researched supposed grain-drying ovens (originals dated to the 4th century AD were found in Hertford). He notes that the general consensus among archaeologists about the purpose of these structures goes back to the 1920s, and it used to be largely uncontested—regular findings of traces, such as carbonised seeds, supported this interpretation (Reynolds 1979, 27). Through his experimentation, Reynolds was able to validate a hypothesis that such structures *could have been used* for drying grain (1979, 38). However, in the process of the experiment, he began to doubt whether such drying approach was reasonably efficient or even necessary. "If these structures are not grain driers and in the light of the evidence to date it is improbable, it is necessary to propose an alternative function rather than provide yet another negative" (Reynolds 1979, 41). Given these considerations, Reynolds proposed and later tested an alternative hypothesis – i.e., the structures in question were malting floors. In a later text, Reynolds even provides more background to his reasoning:

> A visiting brewer, in the late 1970s, challenged the interpretation of a structure as a Romano-British grain drying oven, suggesting that it was far better as a malting floor. Further research proved his hypothesis to be far more probable! (Reynolds 1999, 134).

(2) *An experimental testing of a historical hypothesis* can be tested through physical experiments. However, as practitioners themselves stress in a Popperian manner, a successful test provides a comparatively weak warrant for a claim about the past. On the other hand, a negative outcome offers a much stronger warrant for a claim regarding the lives of our ancestors. Another aspect of note is that (2) *experimental historical hypothesis* does not help identify any new evidence in the form of past traces like artefacts or texts. It builds upon previously accepted evidence and provides warrants for historical claims about the past through distinct intellectual and experimental activities.

One aspect of (2) *An experimental testing of a historical hypothesis* archaeologists themselves tend to overlook in their methodological musings is a move from *tokens* to *types* and from *trace-based reaso*ning to *analogous reasoning*. In some cases, archaeologists are not interested in surviving

artefacts as tokens. They do not inquire into causal chains involved in constructing the particular (supposed) grain dryer; they are not interested in exact years or its builders as individuals. They are interested in the structures as instances of *a type*. Instead of a causal downstream history of the particular artefact, they are interested in the purpose analogous structures could have been regularly used for in the past. The trace-based reasoning (i.e., finding carbonised seeds in the structures) suggested a hypothesis, but experimental testing showed significant issues with such an interpretation (lack of efficiency). The analogous application of expert knowledge from other fields (brewing) helped formulate a competing hypothesis that was not previously available. Despite the insistence of experimental archaeologists that a negative outcome of the experiment is generally more decisive, Reynolds accepted his experiment's positive outcome as a sufficient warrant for labelling the structures in question as malting floors instead of corn dryers. A more robust network of inferences helped to overcome Popperian scepticism about the informational value of a successful experiment.

In this way, experimental archaeology truly works as an experimental science, but we can still understand it as producing a warranted statement about the past. The artificial structures were most likely used as malting floors. This is a knowledge of the past that was not available to archaeologists in the 1920s and led to the mislabelling of actual traces of the past on the basis of crude analogous reasoning. Future traces of the same type that are yet to be discovered will thus inform us about the production capabilities of a particular (token) settlement in a given location. We may hence agree with Ben Jeffares, who argues for blurring the lines between experimental and historical sciences:

> The best way to understand the historical sciences is to see them deploying well understood regularities, particular process types, across multiple tokens, either as a means to secure relationships with evidence, or as a general pattern of explanation (Jeffares 2008, 475).

4. Information-Enriching Hypothesis Testing

Well-understood regularities are important for archaeology in general. Let us briefly explore a recent discovery of an artefact with substantial potential for changing established historical discourse. Contemporary historical accounts of Early Slavs in Central Europe are preoccupied with tracing their migration. In the present-day Czech Republic, historical consensus supposes that there was no direct contact between Slavs, arriving after 556 AD, and Germanic tribes (Lombards), who disappeared in approx. 568 AD. The idea that Slavs in the area were not originally influenced by Germanic tribes is a part of accepted historical narratives (e.g., accepted by a community of historians, taught in schools, etc.; see Chlup 2020), and there used to be no evidence to the contrary. However, among the artefacts found at an Early Slavic settlement near Lány, archaeologists have recently discovered a bone fragment bearing a runic inscription (part of *fuþark*). In the context of established historical knowledge, this artefact has significant potential to provide a warrant for a different narrative. Archaeologists, led by Jiří Macháček (Macháček et al. 2021), have recognised this potential due to their professional training, but in order for evidence to move the needle, additional steps were required. To test their hypothesis, they had to employ various techniques to examine the artefact. Through radiocarbon dating, aDNA analysis of animal bones, taxonomical enrichment, runology, etc., the archaeologists have dated the inner bone section to 585–640 AD. Furthermore, they concluded that the inscription was carved by an inexperienced person who was probably practising the runic alphabet; thus supporting the hypothesis that the artefact is evidence for some kind of contact between two ethnolinguistic groups. The exact nature of this contact remains significantly underdetermined (Germanic tribe members living among Slavs, trade exchange, spoils of war, etc.); however, the successful "test of the hypothesis" (i.e., warranted inquiry into the properties of an artefact) increased the strength of evidence, which now warrants a theory going against the accepted historical discourse and can be used as a strong claim in an argumentative historical practice regarding Early Slavic migration in Central Europe.[5]

[5] It should be noted that professional historical discourse may still raise some objections regarding the finding, e.g., Florin Curta (2009) states that it might be

On the one hand, the testing of the hypothesis required complex technological processes that were facilitated by broad international cooperation. On the other hand, hypothesis testing was straightforward on a theoretical level since it involved (3) *information-enriching hypothesis testing.* Archaeologists do not run the aforementioned laboratory tests without a hypothesis—i.e., without a theoretical suspicion that the artefact may carry some crucial information. If the artefact had been shown to be fake or of a much later date, it would not have been able to function as evidence for a claim concerning the unprecedented contact of two ethnolinguistic groups. The hypothesis would have been adequately infirmed. The successful test of the hypothesis establishes the importance and uniqueness of evidence (i.e., its status as evidence for a certain theory is established), thus achieving a significant milestone in historical discourse. We might say that the theory "is not so much tested as it is used to enhance the informational value of the evidence" (Kosso and Kosso 1995, 591).

Other examples of testing (3) *information-enriching hypothesis testing* may include dating documents, assessing authorship of texts through mathematical modelling, inquiring into chemical properties in the food history context, analysing structural properties of buildings or tools, etc.

Here, we can see that a hypothesis regarding a trace's informational value *and the potential impact this information can have on a professional discourse* is necessary to justify complex and collaborative testing that enriches the information we can obtain from a particular trace. Without such hope for impactful results, historians and archaeologists do not dedicate their time and resources to every trace they have in their inventory. As such, a number of background theories and inferences are usually hidden from the broader public, but these considerations and the rules of the game of giving reasons play a crucial role in historical discourse. Epistemic diligence dictates that token evidence must undergo significant scrutiny before it can substantiate a claim that undermines previous theories about the past. The information obtained in (3) *information-enriching hypothesis testing* may turn a trace into evidence for a novel and groundbreaking claim about the past, as was the case with runic bone from Lány; or it might

misleading to associate specific material culture with specific linguistic developments.

disqualify a trace as evidence for such claim, e.g., in the case in which the evidence is proven to be of much later date or to be a forgery.

5. Contextualising Historical Hypothesis Testing

Let us now explore one last type of historical hypothesis in this paper. We may encounter explicit hypotheses in another subfield of historical disciplines – namely in intellectual history and related subfields like the history of philosophy. In these cases, historians strive to achieve a better understanding of historical agents and their thoughts or utterances. It may seem that hypotheses and testing have no place in this context. However, when analysing the thoughts of our ancestors, historians often make very explicit hypotheses that serve as tools to contextualise evidence (i.e., text, fragment, etc.). The most explicit use of this approach in a theoretical reflection could be found in the works of Quentin Skinner. Influenced by Collingwood (Skinner 2002, 115), he sees the goal of intellectual history as a recovery of past agent's intentions. To do so without subscribing to Collingwood's idealism, he relies on the speech act theory of John L. Austin (2002, 133). The texts that intellectual historians examine are instances of speech acts with particular illocutionary force (2002, 109). To decode the intention of the past agent, we need to know the context of the speech act. However, the context itself must be established first via historical research, and there are often several contenders for the context historians can utilise while following Skinner's methodological proposition. It might be tempting to say that the context is co-determined by a past actor's intention, but this leads to circularity, as Skinner himself acknowledged in an interview:

> I would say that the context is whatever you need to reconstruct in order to understand some meaningful item in that context. This is circular, of course, but I am speaking of a hermeneutic circle. You need to think of texts as answers to questions, and the context as the source of the questions (Li 2016, 122).

Thus, intellectual historians are invited to "test out" different contexts and provide rational argumentation for their choices in order to understand specific utterances. Contexts intellectual historians use are a result of historical

research, and they are often expressed in the form of widely accepted theories within a historical discourse. However, in some cases, intellectual historians might engage in more experimental behaviour.

For instance, exploring ancient Greek philosophy poses multiple challenges to historians. Apart from vast temporal and cultural distance, the often fragmentary nature of texts and gaps in historical records make reaching a consensus about ancient philosophers problematic. A perfect example is Heraclitus of Ephesus' fragmentary and notoriously obfuscated philosophy. Fragments of his work survived through secondary sources, and their reliability and authenticity are often questionable. In his attempt to reconstruct Heraclitus' ethics, David Sider (2013) explores several fragments and tries to contextualise them to achieve a better understanding. A significant portion of this endeavour consists in exploring fragment B29 ("The best choose one thing above all, the everlasting fame of mortals; the many gorge themselves like cattle") as an allusion to Simonides' lyrics dedicated to the fallen at Thermopylae ("Stranger, bring the message to the Spartans that here we remain, obedient to their orders."). The supposed allusion is based on minor linguistic similarities.

Sider acknowledges that the idea itself is a problematic hypothesis in the context of historical discourse regarding Heraclitus because the relative chronology between Heraclitus and Simonides is underdetermined:

> The ancient testimony for the death of the former is rather confused, but 484 would seem to be the absolute earliest date; a later date remains quite possible. If so—and this is what I believe to be the case—it seems to more likely that Heraclitus was responding to Simonides (and the favorable reaction his poem no doubt received in Ephesus) than the reverse (Sider 2013, 326–327).

Sider references historical discourse and various critical discussions regarding the year of Heraclitus' death, which makes his hypothesis less probable. At the same time, it is known that Heraclitus often referenced and called out his contemporaries and predecessors, which lends credence to the hypothesis. There is no imaginable method for historians to test this hypothesis unless some supporting evidence is found by pure chance, i.e., this type of hypothesis does not point towards any archive or place where to look. The hypothesis merely contextualises the fragment as a part of a larger

discourse, and its potential interpretative merits are tested only via intellectual reconstruction. In this case, Sider uses the hypothesis to construct an appealing interpretation of Heraclitus' ethics that is consistent with the complex theories of Plato (Sider 2013, 334). Nonetheless, it is questionable whether this "test" (arriving at a plausible and coherent interpretation) provides a solid warrant for the pivotal claim that Heraclitus had been able to allude to Simonides' poem before his death.

In other cases, debates among historians of philosophy may concern, e.g., the relative importance of specific statements in the historical sources as they may contradict other statements in the broader intellectual context of the historical agent's work. A more recent example could be the so-called New Hume debate and the question of whether David Hume was a realist regarding the necessary connexion (see, e.g., Peterková 2015).

We may call it an *interpretative historical hypothesis testing* or (4) *contextualising historical hypothesis testing*. This type of hypothesis is common in subfields like intellectual history, history of political thought, or history of philosophy. It is concerned with interpreting and understanding historical texts, and it is often more explicit than the previous (1) *evidence-seeking hypothesis testing*. It hypothetically holds some statements about the past to be true, connects them to examined texts or thoughts, and the test consists of evaluating resulting interpretations. Even though (4) *contextualising historical hypothesis testing* might be explicitly formulated, the process of testing is not straightforward, and it often consists of historians claiming that their contextualisation provides better results than contesting interpretations in historical discourse. It might be said that we are stretching the meaning of hypothesis testing too far. Yet, these thought operations and related inferences are important in the discourse of intellectual history. The exact nature of this argumentative process might pose a fascinating topic for philosophers of historiography who may study its rules and results.

When (4) *contextualising historical hypothesis testing* allows for meaningful understanding of historical texts (e.g., Sider's daring hypothesis about relative chronology of Heraclitus and Simonides allows for meaningful reconstruction of Heraclitus' ethics), it may be considered as a positive result; however, it does not warrant rejecting other viable interpretations. Contrary to that, when a hypothesis produces incoherent or confusing results, it gains no

support from the process, and it is likely to be rejected by the professional discourse. The evaluation of (4) *contextualising historical hypotheses testing* takes place in an argumentative field of historical discourse.

5. Conclusion

We have explored four types of hypotheses that historians and archaeologists make and how they test them in order to make warranted claims about the past in the context of argumentative historical discourse. This list does not aspire to be exhaustive, and other types of hypothesis testing might be identified further across various historical disciplines.

(1) *The evidence-seeking hypothesis testing*
(2) *The experimental historical hypothesis testing*
(3) *Information-enriching hypothesis testing*
(4) *The contextualising historical hypothesis testing*

Type (2) is the most typical example of hypothesis testing. The hypothesis must be explicitly worded, and its testing follows strict rules and compares two sets of data. On the surface, its proponents retain strong Popperian principles, and a negative testing result is seen as having a higher informational value than a positive outcome. It relies on *analogous reasoning* and is often concerned with *types*. Hypotheses (3) are needed to explore the evidential potential of historical data that historians or archaeologists identified as promising. Without a hypothesis about the past, complex and expensive laboratory tests are not warranted for every artefact or text. A positive testing of such a hypothesis introduces a new piece of evidence into historical discourse. Negative results of such testing usually do not move the field further.[6] We do not usually see explicit wording of type (1) hypothesis testing in publications of historians; however, it is an integral part of historians' legwork. The skill to make such hypotheses is crucial for historical practice. By analysing historical discourse and available evidence, a

[6] However, when a historical record is identified as fake, a different set of historical questions might be asked. Who created the forgery and what was the intention behind it? Exploring these options may produce different hypotheses.

historian hypothesises about the past and potential traces that might be preserved at a specific place (archive, library, etc.). As such, inquiries can be made to test this hypothesis, and the result can be presented in an argumentative context of historical discourse or analysed via Bayesian framework. Finding a relevant document (either confirming or refuting the hypothesis) can be seen as a milestone and may open new avenues for historical discourse. In contrast, failure to obtain relevant data only infirms the hypothesis and does not further historical discourse. Type (4) happens in an argumentative context, does not concern new evidence, and is mostly derived from historical discourse. A hypothesis is often explicitly stated (i.e., it is argued that some context might be relevant, some configuration of past events is presupposed) and tested only via intellectual reconstruction and interpretation. The results of this experimental testing of different contexts are quite surprisingly similar to type (2). A successful (4) *contextualising historical hypothesis* may produce a viable and enriching interpretation, but it may co-exist with other contending interpretations. A failure to produce a coherent interpretation can be seen as a refutation of a hypothesis.

By exploring these selected four types of historical hypotheses, we may see that any excursion into the argumentative context of professional historians as envisioned by Kuukkanen might benefit from going beyond historical narratives and exploring what Goldstein termed the infrastructure of history and which has a long tradition in the philosophy of historical sciences. We may see that the four sketched types of hypotheses testing exhibit different logical structures, different types of reasoning (trace-based, analogous), and some unexpected similarities. Further instances of hypothesis testing can be identified, e.g., Peter Turchin's cliodynamics, which proceeds via testing different hypotheses concerning social and developmental dynamics in history against vast digital databanks of archaeological and historical records (Turchin et al. 2023).

One lesson that should be considered from the present exercise is that if philosophers of history argue for the practice revolt and closer examination of historical discourse and associated argumentative practices, they need to engage with historical discourse on a much deeper level than just on the level of large synthesising pieces of historical literature. Any warranted statement about the past is part of a complex network of inferences that

are difficult to evaluate from reading a single piece of historiography. I do believe that Kuukkanen would not object to this claim. However, so far, it seems that the postnarrativist project is still wedded to the narrativist tenets of exploring historiography as a practice of history writing, and the vantage point does not allow for exploring the infrastructure of history. Philosophers of historical sciences have already shown the potential of exploring the actual practice of historical scientists, and philosophers of history may benefit from closer cooperation. It does not mean that historical narratives and their rhetorical dimensions should be overlooked but that they should not entirely overshadow the underlying infrastructure of history, consisting of evidential, trace-based, and analogous reasoning, informational epistemology, and hypothesis testing.

Funding

This article and research were funded by GAČR: GA23-05800S – "The Big History and its Philosophical Potential."

Acknowledgements

I am very grateful to the philosophers who have commented on this paper in various stages of production, most notably Aviezer Tucker, Jouni-Matti Kukkanen, Georg Gangl, Eugen Zeleňák, and Adam Timmins. I am also indebted to anonymous reviewers for their constructive comments.

References

Chapman, Robert and Alison Wylie. 2016. *Evidential Reasoning in Archaeology*. London: Bloomsbury Academic.

Chlup, Radek. 2020. "Competing myths of Czech Identity." *New Perspectives*, 28: 179–204. https://doi.org/10.1177/2336825X20911817

Cleland, Carol E. 2001. "Historical Science, Experimental Science, and the Scientific Method." *Geology*, 29: 987–90. https://doi.org/10.1130/0091-7613(2001)029%3C0987:HSESAT%3E2.0.CO;2

Cleland, Carol E. 2013. "Common Cause Explanation and the Search for a Smoking Gun." *Special Paper of the Geological Society of America*, 502: 1–9. https://doi.org/10.1130/2013.2502(01)

Collingwood, Robin G. 1994. *The Idea of History*, edited by Jan van der Dussen. Oxford: Oxford University Press.

Currie, Adrian. 2018. *Rock, Bone and Ruin*. Cambridge, Massachusetts: The MIT Press.

Curta, Florin. 2009. "The Early Slavs in Bohemia and Moravia: A Response to My Critics." *Archeologické Rozhledy*, 61: 725–54.

Dunlop, Rory and Jōn Vidar Sigurdsson. 1995. "An Interdisciplinary Investigation of Bergen's Forgotten Fire: Confrontation and Reconciliation." *Norwegian Archaeological Review*, 28: 73–92. https://doi.org/10.1080/00293652.1995.9965586

Feder, Kenneth. 2019. *Frauds, Myths, and Mysteries: Science and Pseudoscience in Archaeology*. 10th edition. Oxford: Oxford University Press.

Fritze, Ronald H. 2009. *Invented Knowledge: False History, Fake Science and Pseudo-Religions*. London: Reaktion Books Ltd.

Goldstein, Leon J. 1976. *Historical Knowing*. Austin and London: University of Texas Press.

Goldstein, Leon J. 1962. "Evidence and Events in History." *Philosophy of Science*, 29: 175–94.

Goldstein, Leon J. 1986. "Impediments to Epistemology in the Philosophy of History." *History and Theory*, 25: 82–100.

Gruner, Rolf. 1968. "Historical Facts and the Testing of Hypotheses." *American Philosophical Quarterly*, 5: 124–29.

Jeffares, Ben. 2008 "Testing Times: Regularities in the Historical Sciences." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 39: 469–75. https://doi.org/10.1016/j.shpsc.2008.09.003

Kosso, Peter. 2011. *A Summary of Scientific Method*. Dordrecht: Springer. https://doi.org/10.1007/978-1-4614-1308-0

Kosso, Peter and Kosso, Cynthia. 1995. "Central Place Theory and the Reciprocity between Theory and Evidence." *Philosophy of Science*, 62: 581–98.

Kuukkanen, Jouni-Matti. 2015. *Postnarrativist Philosophy of Historiography*. London: Palgrave Macmillan. https://doi.org/10.1057/9781137409874

Kuukkanen, Jouni-Matti. 2017. "Moving Deeper into Rational Pragmatism." *Journal of the Philosophy of History*, 11: 83–118. https://doi.org/10.1163/18722636-12341362

Kuukkanen, Jouni-Matti. 2021. "Historiographical Knowledge as Claiming Correctly." In *Philosophy of History: Twenty-First-Century Perspectives*, edited by J.-M. Kuukkanen, 44–65. London: Bloomsbury Academic.

Li, Hansong. 2016. "Ideas in Context: Conversation with Quentin Skinner." *Chicago Journal of History*. 2: 119–28.

Macháček, Jiří et al. 2021. "Runes from Lány (Czech Republic)—The Oldest Inscription among Slavs: A New Standard for Multidisciplinary Analysis of Runic Bones." *Journal of Archaeological Science*, 127: 1–8. https://doi.org/10.1016/j.jas.2021.105333

McGrew, Timothy. 2014. "The Argument from Silence." *Acta Analytica*, 29: 215–28. https://doi.org/10.1007/s12136-013-0205-5

Murphey, Murray G. 2009. *Truth and History*. New York: State University of New York Press.

Outram, Alan. 2008. "Introduction to Experimental Archaeology." *World Archaeology*, 40: 1–6.

Peterková, Eva. 2015. "David Hume and So-Called 'New Hume.'" *Perspektywy kultury*, 13: 201–16.

Reynolds, Peter J. 1999. "Buster Ancient Farm." In *The Constructed Past: Experimental Archaeology, Education, and the Public*, edited by P. Stone, and Philippe Planel, 124–35. London: Routledge.

Reynolds, Peter J. and J. K. Langley. 1979. "Romano-British Corn-Drying Oven: An Experiment." *Archaeology Journal*, 136: 27–42. https://doi.org/doi:10.6067/XCV8JH3Q14

Sider, David. 2013. "Heraclitus' Ethics." In *Doctrine and Doxography: Studies on Heraclitus and Pythagoras*, edited by D. Sider and Dirk Obbink, 321–34. Berlin: De Gruyter.

Skinner, Quentin. 1969. "Meaning and Understanding in the History of Ideas." *History and Theory*, 8: 3–53.

Skinner, Quentin. 2002. *Visions of Politics*. Volume I: *Regarding Method*. Cambridge: Cambridge University Press.

Sober, Elliot. 2009. "Absence of Evidence and Evidence of Absence: Evidential Transitivity in Connection with Fossils, Fishing, Fine-Tuning, and Firing Squads." *Philosophical Studies*, 143: 63–90. https://doi.org/10.1007/s11098-008-9315-0

Tucker, Aviezer. 2004. *Our Knowledge of the Past: A Philosophy of Historiography*. Cambridge: Cambridge University Press.

Tucker, Aivezer. 2011. "Historical Science, Over- and Underdetermined: A Study of Darwin's Inference of Origins." *British Journal for the Philosophy of Science*, 62: 805–29. https://doi.org/10.1093/bjps/axr012

Tucker, Aviezer. 2025. *Historiographic Reasoning*. Cambridge: Cambridge University Press.

Turchin, Peter, et al. 2023. "Explaining the Rise of Moralizing Religions: A Test of Competing Hypotheses Using the Seshat Databank." *Religion, Brain and Behavior*, 13: 167–94. https://doi.org/10.1080/2153599X.2022.2065345

Turner, Derek. 2007. *Making Prehistory*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511487385

Turner, Derek. 2013. "Historical Geology: Methodology and Metaphysics." *Special Paper of the Geological Society of America* 502: 11–18. https://doi.org/10.1130/2013.2502(02)

Organon F 32 (2) 2025: 193–223 ISSN 2585-7150 (online)
https://doi.org/10.31577/orgf.2025.32204

RESEARCH ARTICLE

Consequences of Explanationism

Nathan William Davies*

Received: 13 June 2024 / Revised: 22 February 2025 / Accepted: 8 March 2025

Abstract: If Explanationism is true, then, for any fact and any person, the fact's explaining why they believe it is both a necessary and a sufficient condition for their knowing it. In this paper I question the sufficiency of the condition and argue against its necessity.

Keywords: Knowledge; belief; explanation; Gettier; skepticism; deduction.

§ 1. Introduction

Bogardus & Perrin (2022; 2023) have recently argued for and defended Explanationism:

> **Explanationism**
> S knows that φ if and only if the fact that φ explains why S believes that φ.

If Explanationism is true, then for any person and any fact, the fact's explaining why the person believes it is both a necessary and a sufficient condition for their knowing it. I refer only to *this* condition with 'the condition', *sans* qualification, in the rest of this paper: if the condition is insufficient,

* Kenilworth
https://orcid.org/0000-0003-4636-9201
Kenilworth, United Kingdom
nwd.correspondence@gmail.com

then in order for someone to know a fact, it is insufficient that it explain why they believe it; if the condition is unnecessary, then in order for someone to know a fact, it is not necessary that it explain why they believe it.

Recently, in the spirit of (Gettier 1963), apparent counterexamples to the sufficiency of the condition have been presented (Piccinini 2022, 411; Mortini 2022, 6–7; Boyce & Moon 2023, 293). Bogardus & Perrin have argued that these apparent counterexamples are not real counterexamples (Bogardus & Perrin 2023). I find Bogardus & Perrin's reply to Mortini's 'defective clock' case adequate (Mortini 2022, 6–7; Bogardus & Perrin 2023, 6–7). However, I think Bogardus & Perrin's reply to Piccinini's case and their reply to Boyce & Moon's case should be discussed.

In § 3 I discuss possible counterexamples to the sufficiency of the condition. In § 3.1 I argue that Explanationists should give positive knowledge verdicts on "fake barn" cases if they reject the safety condition on knowledge. In § 3.2 I discuss Boyce & Moon's 'holoprojector' case and argue that, with respect to a similar case, an Explanationist must attribute knowledge. In § 3.3 I discuss Piccinini's 'telephone' case and Bogardus & Perrin's response to it.

In § 4 I discuss possible counterexamples to the necessity of the condition. In § 4.1 I argue that there are skeptical consequences of Explanationism: if the condition is necessary, then there are cases of apparent knowledge which Explanationism rules out. In § 4.2 I present two counterexamples to the necessity of the condition which both concern deductive knowledge, and I discuss Bogardus & Perrin's 'seeing' account of deductive knowledge.

In the next section, § 2, I discuss Bogardus & Perrin's commitments in detail. I recommend for a first reading that this section be read last after reading the rest of the paper.

§ 2. Complications

§ 2.1 Bogardus & Perrin vs. Strevens

Consider the following passage from (Bogardus & Perrin 2022):

> According to Explanationism…knowledge requires…that beliefs bear the right sort of explanatory relation to the truth. In slogan

> form: knowledge is believing something *because* it's true. Less roughly, Explanationism says knowledge requires…that truth play *a crucial role* in the explanation of your belief. What's a crucial role? We propose adapting Michael Strevens' (2011) "kairetic test" for difference-making in scientific explanation, to the more general purposes of Explanationism. With regard to scientific explanation, Strevens' proposal is this: Start with a deductive argument, with premises correctly representing some set of influences (potential explanans), and the conclusion correctly representing the explanandum. Make this argument as abstract as possible, while preserving the validity of the inference from premises to conclusion; strip away, as it were, unnecessary information in the premises. When further abstraction would compromise the validity of the inference, stop the process. What's left in the premises are difference-makers, factors that play a "crucial role" in the scientific explanation. Now, explaining why a belief is held is importantly different from paradigm cases of scientific explanation. In order to adapt the kairetic test to Explanationism, we propose beginning with a set of potential explanans that explain the relevant beliefs being held, and then proceeding with the abstraction process until further abstraction would make the explanation fail. The remaining explanans are the difference-makers. On Explanationism, for the belief to count as knowledge, the truth of the relevant belief must be among these difference-makers... (Bogardus & Perrin 2022, 179).

Even though Bogardus & Perrin appeal to Strevens' account of difference-making, there is at least one difference between Bogardus & Perrin's account and Strevens' account which is of importance.

In the passage quoted above Bogardus & Perrin describe an abstractive procedure for identifying what Strevens calls an *explanatory kernel*, or, simply, a *kernel*—a set of jointly sufficient explanantia for an explanandum which satisfies certain conditions (see, e.g., (Strevens 2011, 110)) for these conditions). But for Strevens there can be (and there typically is) more than one kernel for a given fact (Strevens 2011, 91, 117–119). So, there can be difference-makers for a fact which are part of one kernel and not another. Hence identifying a kernel using the abstractive procedure is (typically)

insufficient for determining all the difference-makers. So, it is not the case, contrary to what Bogardus & Perrin say, that once you have finished with the abstractive procedure "[t]he remaining explanans are *the* difference-makers" (my emphasis); while the remaining explanantia will be difference-makers, (typically) there will be at least one difference-maker which is not one of those explanantia. For Strevens, a *difference-maker* for a fact is a fact which is part of *at least one* kernel for it (Strevens 2011, 87–88).

This is important for at least one reason: suppose a fact, y, and some other facts, cc, are a kernel for z; if the abstractive procedure sufficed to determine all the difference-makers, we could deduce, for any explanans of y which is not one of cc, call it x, that it could not be a difference-maker for z; for a would-be kernel involving x, y, and cc would always be too complex to be a real kernel, given that y and cc are alone a kernel and x is unnecessary to deduce z given y and cc. But could there not be cases in which: (i) a fact x is a difference-maker for fact y which is in turn a difference-maker for fact z; and (ii) x is a difference-maker for z?

Accounting for cases of this kind is necessary if Bogardus & Perrin are to give an account of knowledge by appealing to difference-making. In an attempt to deal with the following example (which is of the aforementioned kind) they end up appealing to things like the "completeness" and "satisfactoriness" of an explanation in spite of the fact that these things play no part in the abstractive procedure they describe (which was supposed to suffice for determining the difference-makers):

> Consider an ordinary case [of perceptual belief], for example the belief that there's a computer in front of you, formed on the basis of visual perception under normal conditions. Your believing this is straightforwardly explained by the fact that there is a computer in front of you. It's true that there's more we could (and perhaps even should) add to the explanation —facts about the lighting, distance, functioning of your visual system, your visual experience, etc.—but the truth of your belief would remain a crucial part of this explanation. If we tried to end the explanation like so, "You believe there's a computer before you because it looks that way," we'd be ending the explanation on a cliffhanger, as it were. We should wonder whether things look that way because they are that way, or for some other reason. A satisfying,

> complete explanation in this case, then, will include the fact that there is a computer before you, which is the truth of the relevant belief. So, Explanationism tells us this is a case of knowledge, as it should (Bogardus & Perrin 2022, 186).

It is plausible that there is a kernel for the fact that you believe there is a computer in front of you which includes the fact that it looks to you as though there is a computer in front of you and which does not include the fact that there is a computer in front of you. If so, you might, following Bogardus & Perrin's description of the abstractive procedure, wrongly infer that the fact that there is a computer in front of you is not a difference-maker for the fact that you believe it. The fact that Bogardus & Perrin themselves appeal to "satisfactoriness" or "completeness" in this case is evidence that they themselves recognize that their prior description of the identification of the difference-makers cannot be quite right.

What Bogardus & Perrin must do, if they are to follow Strevens at all, is to stick to Strevens' account more closely: as stated before, for Strevens a fact x is a difference-maker for a fact y if and only if x is part of *at least one* kernel for y. Strevens shows that given his understanding of difference-makers and kernels: if a fact x is part of a kernel for y, and y is part of a kernel for z, then x is part of a kernel for z (Strevens 2011, 119, 122); and if a fact, x, is a difference-maker for a fact, y, and y is a difference-maker for z, then x is a difference-maker for z (Strevens 2011, 121–122). The following consequence is just what Bogardus & Perrin want: as long as the fact that there is a computer in front of you is a difference-maker for the fact that it looks to you as though there is a computer in front of you, and the fact that it looks to you as though there is a computer in front of you is a difference-maker for the fact that you believe that there is a computer in front of you, the fact that there is a computer in front of you will be a difference-maker for the fact that you believe that there is a computer in front of you. (It is worth stressing here, in response to comments from a reviewer, that Bogardus & Perrin's account of explanation, if it is to follow Strevens' at all and account for cases like the one just mentioned, will entail that explanation, itself grounded in difference-making, is transitive. So Bogardus & Perrin must accept key inferences in the arguments presented in § 3.)

There is one last thing to say about Bogardus & Perrin's account of explanation *vis-à-vis* Strevens': from the description of their procedure, it is not clear what Bogardus & Perrin think about the explanation of undetermined facts. Suppose that no group of explanantia are jointly sufficient for an explanandum, suppose, e.g., that a particular sample of Uranium-238 emitted an alpha particle yesterday and that it was truly undetermined that it would do so. There cannot be a deductive argument of the kind Bogardus & Perrin need there to be to identify the difference-making facts (i.e. a certain sort of deductive argument the premises of which are explanantia of the conclusion), so at best they have no account of undetermined facts and at worst they must conclude that all undetermined facts are inexplicable. Strevens thinks (2011, 359–361) that one can either try to explain the undetermined fact (e.g. the fact that the sample emitted the alpha particle yesterday) and in place of deductive validity have instead some sort of 'probabilistic validity' (Strevens 2011, 359) or one can give up trying to explain the undetermined fact itself and explain instead the associated probability fact (e.g. the fact that at 00:00 yesterday the probability of the sample emitting the alpha particle by 00:00 today was n), keeping the kernel-identification procedure the same as before. These issues matter. For instance, if it is possible for someone to believe something, know it, and yet it have been undetermined that they would believe it, then Bogardus & Perrin must go with the 'probabilistic validity' option. I shall say no more about these issues here.

*§ 2.2 Explanationism**

A reader might justifiably wonder whether Bogardus & Perrin are actually defending Explanationism* rather than Explanationism:[1]

[1] Bogardus & Perrin use "playing a crucial role"-phrases and "making a difference"—phrases interchangeably in (Bogardus & Perrin 2022) and (Bogardus & Perrin 2023), and on this basis I take Bogardus & Perrin to be committed to the following proposition: the fact that φ is a difference-maker for the fact that ψ if and only if the fact that φ plays a crucial role in an explanation of why ψ. There is thus no need to consider a possible third thesis of theirs phrased in terms of "playing a crucial role."

Explanationism*
S knows that φ if and only if the fact that φ is a difference-maker for the fact that S believes that φ.

They might also wonder what Bogardus & Perrin take the relations between Explanationism and Explanationism* to be. It is obvious that DM→E is true (recall that difference-makers are, by definition, explanantia):

DM→E
If the fact that φ is a difference-maker for the fact that ψ, then the fact that φ explains why ψ.

so, it is obvious that Bogardus & Perrin would think it true. I think that Bogardus & Perrin would also think that E→DM and (hence) DM↔E are true too:

E→DM
If the fact that φ explains why ψ, then the fact that φ is a difference-maker for the fact that ψ.

DM↔E
The fact that φ is a difference-maker for the fact that ψ if and only if the fact that φ explains why ψ.

For in my assessment, Bogardus & Perrin speak of "difference-makers," "crucial roles," and Strevens' "kairetic test" to try to separate the real explanantia for a given fact from the merely apparent explanantia; they do *not* speak of such things in order to separate the "difference-making" explanantia from the non-"difference-making" explanantia. Obviously, if DM↔E is true, Explanationism* is true if and only if Explanationism is true.

I will proceed under the assumption that Bogardus & Perrin are interested in defending Explanationism primarily. I shall thus not speak of difference-making in what follows (except in footnotes) and continue with a direct assessment of Explanationism. That said, I try hard not to assume, in what follows, any explanatory facts which could not plausibly be the case if explanation was grounded in difference-making. I think, but will not show, that there are analogous arguments to those given in § 3 which concern difference-making and which are no less plausible. (Note

that my arguments against the necessity of the (Explanationist) condition, given in § 4, also tell against the necessity of the Explanationist* condition given the truth of DM→E.)

§ 3. Is the Condition Insufficient?

3.1 "Fake Barn" Cases

Consider Ginet's (Goldman 2009, 79, 79 fn.5) "fake barn" case:

> Henry is driving in the countryside with his son. For the boy's edification Henry identifies various objects on the landscape as they come into view. "That's a cow," says Henry, "That's a tractor," "That's a silo," "That's a barn," etc. Henry has no doubt about the identity of these objects; in particular, he has no doubt that the last-mentioned object is a barn, which indeed it is. Each of the identified objects has features characteristic of its type. Moreover, each object is fully in view, Henry has excellent eyesight, and he has enough time to look at them reasonably carefully, since there is little traffic to distract him.…[U]nknown to Henry, the district he has just entered is full of papier-mâché facsimiles of barns. These facsimiles look from the road exactly like barns, but are really just facades, without back walls or interiors, quite incapable of being used as barns. They are so cleverly constructed that travelers invariably mistake them for barns. Having just entered the district, Henry has not encountered any facsimiles; the object he sees is a genuine barn. But if the object on that site were a facsimile, Henry would mistake it for a barn (Goldman 1976, 772–773).

What should Explanationists say about "fake barn" cases such as this? I think that they should affirm, for any "fake barn" case, that the relevant person (e.g. Henry) *does* know that there is a barn in front of them and Explanationists should resist the temptation to deny that the person believes that there is a barn in front of them because there is a barn in front of them (regardless of how many fake barns there are, where they are positioned, etc.). They should also give the same verdict on similar cases

involving fake diamonds, fake antelopes, fake tomatoes, boys who cry "wolf" etc. (Carrier 1976, 247–250; Gendler & Hawthorne 2005).[2]

Bogardus & Perrin seem reluctant to give flat-footed positive verdicts on "fake barn" cases. They write: "We consider it a virtue of Explanationism that its verdicts on Fake Barn Country cases seem to be as indeterminate as our intuitions about whether there's knowledge in such cases" (Bogardus & Perrin 2022, 193, note 24).

The only reason I can think of for an Explanationist not giving a positive verdict on a "fake barn" case is that: (i) she accepts a safety condition on why-explanation:

> **Safety of Why-Explanation**
> If the fact that φ explains why ψ, then, if Ψ, Φ
> (E.g. "If the fact that the owl hit the window explains why I am awake, then, if I were to be awake, the owl would have hit the window," "If …, then I couldn't (easily) have been awake without the owl having hit the window," "If …, then before the owl hit the window it was much more likely for me to be awake and the owl to have hit the window than it was for me to be awake and the owl not to have hit the window.")

and (ii) she judges that the relevant person in the "fake barn" case could easily have believed that there was a barn in front of them without there being a barn in front of them. But Bogardus & Perrin give reasons for rejecting a safety condition on *knowledge* ((Bogardus & Perrin 2022, 181–182); (Bogardus 2014)). And given that an Explanationist should reject a safety condition on knowledge, they should reject the Safety of Why-Explanation; for Explanationism and the Safety of Why-Explanation jointly entail the Safety of Knowledge:

α

1 If S knows that φ, then the fact that φ explains why S believes that φ.
[Explanationism (the necessity of the condition)]

[2] And if Bogardus & Perrin are right that Plantinga's "brain lesion" case (Plantinga 1988, 22–33; Plantinga 1993, 195) is basically a "fake barn" case (2022, 192–93), they should give a positive knowledge verdict on that case too.

2 If the fact that φ explains why S believes that φ, then, S couldn't (easily) have believed that φ without Φ.
[Safety of Why-Explanation]

If S knows that φ, then, S couldn't (easily) have believed that φ without Φ.
[1 & 2; Safety of Knowledge]

and indeed Bogardus & Perrin themselves *do* reject the Safety of Why-Explanation (Bogardus & Perrin 2022, 185, note 15).

In spite of their rejection of the Safety of Why-Explanation, Bogardus & Perrin make the following comparison:

> Even if your eyes happen to fall upon a real barn in a forest of fakes, we might begin to think it false that *it looks like there's a barn before you because there is a barn before you*. As the barn facades proliferate, a rival explanation looms into view: that it looks like there's a barn before you because you're in a region full of structures that look like barns. … Compare: While driving, a contractor's truck drops a large number of sharp objects – nails and screws – all over the road. The car following behind the truck gets a flat tire because it ran through this mess. Meditate for a moment on the suitability of this explanation. Now, while it may be true that one particular sharp object – a nail, let's say – punctured the tire, it's unnecessary to cite that particular object, or the fact that it was a nail rather than a screw, in order to explain the puncture, given all the other sharp objects nearby that nail, poised to puncture the tire in its place. All that figures crucially into the explanation of the punctured tire is the prevalence of these sharp objects (Bogardus & Perrin 2022, 193).

I do not understand what the significance is of this comparison given their rejection of the Safety of Why-Explanation. If one rejects the Safety of Why-Explanation, surely one must concede that the fact that the nail punctured the tire *does* explain why the tire is flat, regardless of whether the tire could easily have been flat without the nail having punctured it. After all, the nail caused the tire to be flat[3] and it is the case that if x causes y

[3] For a recent defence of the relative fundamentality of thing-causation over event-causation see (Baron-Schmitt 2024).

to be F (by V-ing y) then y is F because x V-ed y.[4] Likewise, if one rejects the Safety of Why-Explanation, surely one must concede that the fact that there is a barn in front of you *does* explain why it looks to you as though there is a barn in front of you, regardless of whether it could easily have looked to you as though there was a barn in front of you without there being a barn in front of you.

As Bogardus & Perrin acknowledge (2022, 193, note 24; 2023, 4–5, 7–8), "fake-barn" cases are *not obviously* cases in which the relevant person *does not know* the relevant fact (Gendler & Hawthorne 2005; (Colaço et al. 2014; Turri 2016, 762–764; Schellenberg 2018, 210–212), so "fake barn" cases aren't obviously counterexamples to the sufficiency of the condition. Given Explanationism and the falsity of the Safety of Why-Explanation, an Explanationist should give positive knowledge verdicts on "fake barn" cases.

§ 3.2 "Holoprojector" Cases

Boyce & Moon present the following 'holoprojector' case as a counter-example to the sufficiency of the condition:[5]

> **Holoprojector**
> Micha sees what appears to be a vase sitting on a pedestal. As it happens, the pedestal is really a holographic projector, and there is no vase on top of it. Rather, what Micha is seeing is merely a realistic holographic projection. Micha, who is ignorant of these facts, comes to believe there is a vase in front of him. As it turns out, hidden in a hollow compartment within the pedestal, out of sight, is a vase. The setup is such that the pedestal projects a realistic holographic image of whatever is in that

[4] Regarding difference-makers (see § 2): surely the fact that the nail punctured the tire would be part of at least one kernel for the fact that the tire is flat, hence it would count as a difference-maker. Bogardus & Perrin saying "it's unnecessary to cite that particular object" (2022, 193) in order to explain why the tire is flat suggests that they think that if the fact that the nail punctured the tire does not appear in *every* kernel, then it is not a difference-maker; but one could only infer that it is not a difference-maker if it did not appear in *any* kernel (§ 2.1).

[5] It is based on a similar one in (Lehrer & Paxon, Jr, 1969, 234–35), which was, it seems (Lehrer & Paxon, Jr, 1969, 234, note 12), inspired by a discussion of holograms in (Goldman 1967, 359).

> compartment onto its surface, and this explains why Micha sees the image before him (Boyce & Moon 2023, 293).

Boyce & Moon think that Explanationists are committed to the soundness of the following argument:

β

1 Micha believes that there is a vase in front of him because he sees a hologram of a vase in front of him.

2 He sees a hologram of a vase in front of him because there is a hologram of a vase in front of him.

3 There is a hologram of a vase in front of him because there is a vase in front of him.

4 Micha believes that there is a vase in front of him because there is a vase in front of him.
[1–3]

Micha knows that there is a vase in front of him.
[4, Explanationism (the sufficiency of the condition)]

Boyce & Moon think that the premises are true, but they think that the conclusion is false. Hence, they infer that the condition is insufficient.

Bogardus & Perrin deny that Explanationists are committed to the soundness of β. They argue for the falsity of Premise 4 and Premise 3: "it's incidental that the device was constructed so that the object to be projected is adjacent to the projected image itself. And, in that case, the fact that the vase is before Micha does not figure crucially into the explanation of Micha's belief" (2023, 10–11); see also (2023, 10–11 fn.17). It is, I think, not unreasonable to reject Premise 3. Consider the following argument against Premise 3 of β:[6]

γ

1 If the fact that there is a vase in front of Micha explains why there is a hologram of a vase in front of him, then either it

[6] It's not clear exactly what Bogardus & Perrin's argument against Premise 3 of β is (Bogardus & Perrin 2023, 10–11, 10–11, note 17). But I think it *is* clear that their argument is sound only if γ is sound.

explains why Micha is oriented towards the hologram or it explains why the hologram is being projected.

2 The fact that there is a vase in front of Micha does not explain why Micha is oriented towards the hologram.

3 The fact that there is a vase in front of Micha does not explain why the hologram is being projected.

The fact that there is a vase in front of Micha does not explain why there is a hologram of a vase in front of him.
[1, 2, & 3]

While it seems to be true that if there weren't a vase in front of Micha, the hologram wouldn't be being projected, it seems not unreasonable to say that it is not the fact that there is a vase in front of Micha, but rather (something like) the fact that there is a vase within the pedestal, which explains why the hologram is being projected.

But now consider the following case:

Art installation
There is an art installation in a gallery. It is a pitch-black room within which there are three black pedestals which cannot be seen. The viewer stands in the centre, and if they have their back to the door there is a pedestal on their left, one in front of them, and one on their right. There is an arced track on which a vase is shuttled between the pedestals at regular intervals. This cannot be seen and the shuttling cannot be heard. While the vase is not being shuttled the vase is inside one of the three pedestals. There is a night-vision camera which tracks the viewer's position relative to the vase. If the viewer is oriented towards the vase (it is the orientation of their body, not their head or eyes, which is relevant) and the vase is inside one of the three pedestals (rather than being on its way to one of the pedestals), then a hologram of the vase appears above the pedestal. If the viewer is not oriented towards the vase, then even if the vase is inside one of the three pedestals, a hologram of the vase does not appear. If the viewer is oriented towards the vase, but the vase is not inside one of the three pedestals, then no hologram appears then either. If the viewer sees a hologram

of the vase, it looks to them as if there were a real vase being illuminated. Micha enters the room, and with his back to the door faces the pedestal in front of him. The vase happens to be inside the pedestal and the night vision camera correctly determines that the vase is in front of Micha. The hologram appears, Micha sees it, and Micha believes that there is a vase in front of him.

It seems that the Explanationist is committed to the following argument being sound (this is the counterpart of β):

δ

1 Micha believes that there is a vase in front of him because he sees a hologram of a vase in front of him.

2 He sees a hologram of a vase in front of him because there is a hologram of a vase in front of him.

3 There is a hologram of a vase in front of him because there is a vase in front of him.

4 Micha believes that there is a vase in front of him because there is a vase in front of him.
[1–3]

Micha knows that there is a vase in front of him.
[4, Explanationism]

The counterpart of γ, ε, seems to be unsound:[7]

ε

1 If the fact that there is a vase in front of Micha explains why there is a hologram of a vase in front of him, then either it explains why Micha is oriented towards the hologram or it explains why the hologram is being projected.

[7] Just to be absolutely clear, the premises and conclusions of β and γ are about what obtains in Boyce & Moon's 'holoprojector' case; the premises and conclusions of δ and ε are about what obtains in the 'art installation' case. This is why the soundness conditions of β and γ are different from the soundness conditions of δ and ε.

2 The fact that there is a vase in front of Micha does not explain why Micha is oriented towards the hologram.

3 The fact that there is a vase in front of Micha does not explain why the hologram is being projected.
The fact that there is a vase in front of Micha does not explain why there is a hologram of a vase in front of him.
[1, 2, & 3]

Premise 3 of ε seems to be false; it is, at the very least, much less plausible than Premise 3 of γ. There doesn't seem to me to be a good argument against Premise 3 of δ, and given that it and the other premises seem true, I conclude that they are true.[8] Now, does Micha know that there is a vase in front of him in my 'art installation' case? If you think not, you should think that the condition is insufficient.

§ 3.3 "Telephone" Cases

§ 3.3.1 Piccinini's "Telephone" Case

Piccinini presents the following 'telephone' case as a counterexample to the sufficiency of the condition:[9]

> **Telephone**
> [C]onsider a situation in which agent A shares their knowledge that *p* (e.g., "the rat is on the vat") with agent B, agent B mishears and whispers *q* to C (e.g., "the cat is on the mat"), but agent C mishears in the opposite direction and comes to truly believe that *p* (Piccinini 2022, 411).

Piccinini infers that the following proposition is true:

> C believes that *p* because *p*.

[8] Bogardus & Perrin cannot object to the inference from Premises 1–3 of δ to Premise 4 given that their account of explanation, grounded in difference-making, entails that explanation is transitive (see § 2).

[9] I call it the 'telephone' case because it reminds one of the children's game of the same name.

And he thinks that Explanationists, on account of their commitment to the sufficiency of the condition, are thereby committed to C knowing that *p*. But Piccinini thinks that C does not know that *p*. Hence he thinks that the condition is insufficient and Explanationism is false. Bogardus & Perrin's reply to this case resists summarization, so I ask the reader's patience as I go through the various stages of it.

§ 3.3.2 Bogardus & Perrin's first variant

Bogardus & Perrin firstly consider a variant of the "telephone" case in which B and C have complementary dispositions which are such that B and C "always…reverse each other's mistakes" (2023, 5):

> [I]f that is the case, then receiving testimony in this way would in fact be a reliable guide to the truth. It would be a bit like a process of photography whereby a scene is transferred onto film with reversed tones as a negative, and then the negative image has its tones reversed again when it is developed, resulting in a faithful representation of the scene. One can know, on the basis of the final, developed film, what the original scene was like, despite this double-reversal of tones involved in the photographic process. If something like that is the case with agents B and C – which, admittedly, would be rather bizarre, even by the standards of philosophical thought experiments – then while the truth of agent C's belief figures into the explanation of why he holds it, [E]xplanationism gives the right result: this is knowledge (Bogardus & Perrin 2023, 5)

It is worth questioning whether this is the right result. This variant of the 'telephone' case is analogous to the following case of dodgy abductive reasoning:[10]

[10] This particular example is due to Maria Zanella. This case has the following form: the fact that φ explains why S believes that φ; S (invalidly) infers that χ because S believes that φ; S believes that χ because S (invalidly) infers that χ; S (invalidly) infers that ψ because S believes that χ; S believes that ψ because S (invalidly) infers that ψ; the fact that ψ explains why φ; hence the fact that ψ explains why S believes that ψ.

Scurvy

> The fact that a man had yellow eyes the last time his mother saw him explains why his mother believes that he had yellow eyes. She infers from the fact that he had yellow eyes that he was a demon. (She thinks that if someone has yellow eyes then they are very likely to be a demon.) She comes to believe that her son was a demon. She infers, from the premise that he was a demon, that he had scurvy. (She thinks, for some odd reason, that all demons have scurvy, and is ignorant of the fact that having yellow eyes is a symptom of having scurvy—she thinks that if you have yellow eyes you are very likely to have scurvy, but she doesn't think that if you have scurvy you are likely to have yellow eyes.) She comes to believe that her son had scurvy. She believes that he had scurvy because she inferred that he had scurvy and because she believed, when she inferred, that he was a demon. She believed that he was a demon because she inferred that he was a demon and because she believed that he had yellow eyes. As stated, she believed that he had yellow eyes because he had yellow eyes. As a matter of fact, he had yellow eyes because he had scurvy. So, she believes that he had scurvy because he had scurvy.

Do Bogardus & Perrin really want to say that the mother in the 'scurvy' case knows that her son had scurvy? If not, then what relevant difference can they point to between the 'scurvy' case and the variant of the 'telephone' case under consideration which justifies the differing verdicts? I can't see any. Does it count in favour of her knowing that her son had scurvy that she would, on account of her beliefs, *reliably* infer that someone had scurvy from their having yellow eyes? I think not. Likewise, I think the fact that B and C would reliably reverse each other's mistakes doesn't count in favour of C knowing that *p*.

§ 3.3.3 Bogardus & Perrin's Second Variant

Bogardus & Perrin secondly consider a variant of the 'telephone' case in which "B and C mishear things quite at random, and there's far from any guarantee that their mistakes will reverse each other" (2023, 5):

> [I]f that's what's happening, then agent C believes that the rat is on the vat not because of the particular content of what he heard from agent B, but because he heard *something or other*, and scrambled it. And agent B said "the cat is on the mat" because that's what he heard, but that's what he heard *not* because agent A said "the rat is on the vat," but rather because agent A said *something or other*, which agent B scrambled. While agent A did say something because he wished to report that a rat is on the vat, the particular content of his belief is not a "difference-maker"[11] in this chain of explanation. It doesn't really matter what exactly he wished to report, only that he initiated this chain of events by reporting something or other. If the scrambling on the part of agents B and C is random, as we're now considering, then it's just a fluke that agent C ended up believing something that matched the original input from agent A. It was just as likely that agent C would have ended up with that belief had agent A said something else entirely. So, agent C's belief that the rat is on the vat is not held because it's true. It's held because agent A initiated a process of random scrambling, which by the sheerest coincidence happened to result in this belief on the part of agent C (Bogardus & Perrin 2023, 5).

I think that what Bogardus & Perrin *actually* say here leaves open the possibility that C knows that p, contrary to what they wish to show. In the first sentence they commit themselves to the truth of the following proposition:

> C believes that p because B said something to C.

But this is the crucial premise of the following argument:

ζ

1 A believed that p because p.

2 A said something to B because A believed that p.

3 B said something to C because A said something to B.

4 C believes that p because B said something to C.

[11] See § 2 for a discussion of difference-making.

5 C believes that p because p.
[1–4]

C knows that p.
[5 & Explanationism]

Premise 1 follows from the necessity of the condition and the stipulation that A knows that p. It is unclear whether Bogardus & Perrin think Premise 2 is true. This is the relevant part of the passage which concerns Premise 2:

> While agent A did say something because he wished to report that a rat is on the vat, the particular content of his belief is not a "difference-maker" in this chain of explanation. It doesn't really matter what exactly he wished to report, only that he initiated this chain of events by reporting something or other (Bogardus & Perrin 2023, 5).

The first clause seems to commit them to the truth of Premise 2,[12] the rest seems contrary to it.[13] They would be wrong to deny Premise 2, however. In the variant of the 'telephone' case under consideration, in which there are 'random' mishearings, it is true that *what* A *said* to B doesn't matter; it is for this reason that we are only considering as part of the explanatory chain the fact that A said *something* to B, and not the fact that A said to B that p. But it does not follow from this or any of the suppositions of the variant under consideration that what A *believed* doesn't matter. It is compatible with the suppositions of the variant under consideration that A said something to B because A believed that p, and that A wouldn't have said anything to B had A not believed that p. So Bogardus & Perrin cannot rule out that Premise 2 is true in the variant under consideration.

Premise 3 is presumably true. Premise 4, as I said, is the crucial premise: given Premises 1–3, if Premise 4 were true, Bogardus & Perrin wouldn't be able to rule out that C knows that p even though this is what they intended

[12] I ignore the fact that Bogardus & Perrin speak of wishing to report something, rather than believing. I don't think this difference is one on which they wish to rest the argument. They even use the expression "his belief" in the second clause as if A's believing has already been mentioned.

[13] It should be clear from my discussion that I don't think the interpretation of the passage hinges on whether Bogardus & Perrin believe E→DM to be true (see § 2.2).

to do. Bogardus & Perrin, in my opinion, *mistakenly* commit themselves to the truth of Premise 4 in the passage quoted above. I shall now explain why I think it a mistake.

If I have understood Bogardus & Perrin correctly, one of the consequences of what they say when they suppose that "B and C mishear things quite at random" is that it was (objectively) unlikely for C to believe that p even at the point at which B said to C that q. And yet if it were objectively unlikely for C to believe that p even at the point at which B said to C that q is it really the case that C believes that p *because* B said something to C? I think that Bogardus & Perrin should say "no." They *should* have said (here I am in fact paraphrasing Bogardus & Perrin (2022, 192)): we may have an explanation of why C believes *something*, but we have no explanation of why C believes that p; that's the nature of a truly random process, there is no demystifying explanation of its operation.[14] In other words, Bogardus & Perrin should say that Premise 4 is false, and hence that ζ is unsound.

This imagined response from Bogardus & Perrin to this variant is, I think, unobjectionable.[15] However I doubt whether discussion of this variant is relevant at all when considering the 'telephone' case of Piccinini: I take it that, unlike the 'random mishearing' variant, in Piccinini's case there is no indeterminacy at all. It is to issues of relevance that I now turn.

§ 3.3.4 Is Bogardus & Perrin's Reply Off-Target?

After considering the two variants Bogardus & Perrin say this:

> There are also possibilities in between. Perhaps agents B and C do not scramble what they hear completely at random, and perhaps they don't reliably reverse each other's errors so as to invariably produce a faithful report in the end. Perhaps they corrupt what they hear only to *some* degree, and in this particular

[14] Bogardus & Perrin should be receptive to this paraphrase given that it is they who refer the reader to the relevant passage of (Bogardus & Perrin 2022) in discussing the "telephone" case (2023, 5).

[15] I set aside the fact that Bogardus & Perrin need a different account of explanation to deal with cases of (objective) indeterminacy (see § 2.1).

> case Piccinini describes it merely *happened* faithfully to preserve the original message. If so, we believe the thing to say is this: the closer the corruption process of agents B and C is to random, the clearer it is that there's no knowledge here, according to [E]xplanationism. And that accords with our intuitions. The closer the corruption process of agents B and C is to the reliable double-reversal process, the clearer it is that there is indeed knowledge here, according to [E]xplanationism. And that too accords with our intuitions. There will be a sliding scale in between these two extremes. But, anywhere you go on that scale, [E]xplanationism delivers verdicts that accord with our intuitions (Bogardus & Perrin 2023, 6).

The first point to make is that I have already called into question whether there is knowledge in Bogardus & Perrin's first variant of Piccinini's case (3.3.2). The second point I want to make is the following. Bogardus & Perrin write as if they have identified *one* axis of variation for cases such as Piccinini's (one "scale"), but it seems to me that they have identified at least two axes pertaining to the following variables: how complementary the dispositions of B and C are (or perhaps simply whether B and C have complementary dispositions); and the likelihood of C coming to believe that p at the time at which A said to B that p (or perhaps at some other time before C's believing that p). There is no *one* scale with cases at one end in which there is no explanation and no knowledge and cases at the other end in which there is both explanation and knowledge. So, if the implication was that the 'telephone' case sits somewhere on this one scale and is accounted for regardless of where it falls, this implication is false.

As I understand Piccinini's "telephone" case, it is a case in which it was determined at the time at which A said to B that p (and indeed at the time at which the rat was on the vat) that C would come to believe that p, but B and C do not have complementary dispositions—there aren't numerous hypothetical 'telephone' circumstances under which they would reverse each other's mishearings. Rather, in Piccinini's 'telephone' case circumstances were favourable and C came to believe a truth. I concur with Piccinini that Explanationists are committed to C knowing that p in this case. If you think that C does not know that p, you should think that the condition is insufficient.

§ 4. The Condition is Unnecessary

§ 4.1 The Skeptical Consequences of Explanationism

In this section, I argue that there are apparent cases of knowledge which, if the condition were necessary, would not be real cases of knowledge. Consider the following 'chute' case based on the one in (Sosa 1999, 145–146):

> **Chute**
>
> On his way to the elevator Smith throws a trash bag down the chute in his apartment building. Smith is now in the elevator. Smith believes that the trash bag is in the basement. The trash bag *is* in the basement.

Does Smith know that the trash bag is in the basement? Not if the condition is necessary. The fact that the trash bag is in the basement does not explain why Smith believes that it is in the basement. So, if the condition is necessary, Smith does not know that the trash bag is in the basement. At best, Smith knows that unless something strange or anomalous happened, the trash bag is in the basement. If you think that Smith *does* know that the trash bag is in the basement, you should think that the condition is unnecessary.[16]

Consider now the following three cases, based on cases in (Starmans & Friedman 2012, 278, 282–283) and (Starmans & Friedman 2020, 14):

> **1936**
>
> Smith's piggy bank contains many coins of different denominations. He doesn't know the dates of any of the coins already in his piggy bank, he doesn't even know *roughly* when they are from. Just before putting a quarter into the piggy bank, Smith uncharacteristically looks at the date on it and sees that it is from 1936. He puts it in the piggy bank, he puts the piggy bank in his closet, and he goes to sleep. He wakes up the next morning, he doesn't check his piggy

[16] I think that Bogardus & Perrin would accept that Smith does not know that the trash bag is in the basement (Bogardus & Perrin 2022, 191, note 21)—the "unless something strange or anomalous" wording comes from their discussion of knowledge of the future.

bank, he doesn't even open the closet, he leaves his room, goes to the kitchen, and is eating his breakfast. The quarter he put in the piggy bank last night has remained there ever since he put it there and it is in the piggy bank right now.

Thief

Smith's piggy bank … and he goes to sleep. While Smith is sleeping, Jones comes into Smith's room, opens Smith's closet, and steals a handful of coins from Smith's piggy bank, including the 1936 quarter Smith put in the piggy bank before he went to sleep. Jones leaves the piggy bank in the closet, closes the closet door, and exits Smith's room. Smith wakes up the next morning, he doesn't check his piggy bank, he doesn't even open the closet, he leaves his room, goes to the kitchen, and is eating his breakfast. The quarter he put in the piggy bank last night is no longer in the piggy bank, Jones has it, but there is another quarter still left in the piggy bank which has been there since before the one Smith put in there last night, which has remained in the piggy bank all night, and which is in the piggy bank right now. Additionally, it is from 1936.

1938

Smith's piggy bank … he doesn't even know *roughly* when they are from. Just before putting a quarter into the piggy bank, Smith uncharacteristically looks at the date on it and comes to believe, based on what he sees, that it is from 1936. As a matter of fact it is slightly damaged and it is from 1938. He puts it in the piggy bank, he puts the piggy bank in his closet, and he goes to sleep. Smith wakes up the next morning, he doesn't check his piggy bank, he doesn't even open the closet, he leaves his room, goes to the kitchen, and is eating his breakfast. The quarter he put in the piggy bank last night has remained there since he put it there and is in the piggy bank right now. There is also another quarter in the piggy bank which *is* from 1936, which has been in the piggy bank since before the quarter he put in there last night, which has remained in the piggy bank all night, and which is in the piggy bank right now.

Now consider the following three questions:

> In the '1936' case, at the time at which Smith is eating his breakfast, does he know that there is a 1936 quarter in his piggy bank?
>
> In the 'thief' case, at the time at which Smith is eating his breakfast, does he know that there is a 1936 quarter in his piggy bank?
>
> In the '1938' case, at the time at which Smith is eating his breakfast, does he know that there is a 1936 quarter in his piggy bank?

If Explanationism is true, and, in particular, if the condition is necessary, then the answer to all three questions is "no." The reason is the same for all three cases: supposing that Smith believes that there is a 1936 quarter in his piggy bank, the fact that there is a 1936 quarter in his piggy bank does not explain why he believes it; he believes that there is a 1936 quarter in his piggy bank either because he believes that there *was* a 1936 quarter in his piggy bank or because he believ*ed* (the evening before) that there was a 1936 quarter in his piggy bank (i.e. he believed "there is a 1936 quarter in my piggy bank").[17]

One might justifiably think that giving the same answer to all three questions is absurd (as opposed to saying "yes," "yes," "no" or "yes," "no," "no") and that it is even more absurd to give the same answer to all three questions for the same reason. If you aren't so skeptical as to give a negative answer to all three questions, you should think that the condition is unnecessary.

§ 4.2 Explanationism and Deduction

§ 4.2.1 Counterexamples

In this section I present two counterexamples to the necessity of the condition. Both of these counterexamples concern deductive knowledge. Before I present my counterexamples, I discuss inferring and deducing.

Inferring(/reasoning) will not be defined. I assume that inferring is thinking (not all thinking is inferring), inferring takes time, and inferring is

[17] These cases raise difficult questions about the accumulation of beliefs over time and about the identity conditions of beliefs over time, questions which I will not try to address here.

a process. I assume that one can believe something (at least in part) because one inferred it. I assume that to believe something is to be in a certain sort of state or else it is to have certain dispositions; I assume that believing something is not a process. I assume that if someone believes something because they inferred it, then for every premise from which they inferred it they believe it/the conclusion (at least in part) because they believed the premise (whether or not the fact that they believed the premise (partially) explains why they inferred the conclusion).

To validly infer will not be defined. However, if one validly infers a conclusion, then it is not possible for the conclusion to be false if the premise(s) one inferred from is(/are) true. I stipulate that to deduce (i.e. soundly infer) something is to validly infer it from facts. Hence if one deduces something it is a fact.

I assume that if someone believes something (at least in part) because they inferred it, and if the inferring was a deducing, and if all of the facts from which they deduced it were known by them at the time of the deduction, then they know it. I assume that if even one of the facts from which they deduced it was merely believed and not known by them at the time of the deduction, then, unless they believe the conclusion for some reason entirely disconnected from the inference, they do not know it.

I think that the following states of affairs are compossible: someone knows a fact; they believe the fact because they inferred it; the fact doesn't explain why they believe it; the inferring was a deducing; and all of the facts from which they deduced the conclusion (i.e. the fact) were known by them at the time of the deduction. If these states of affairs *are* compossible, then the condition is not necessary and Explanationism is false. In the rest of this section, I present two counterexamples to the necessity of the condition which are supposed to illustrate the compossibility of the aforementioned states of affairs.

Consider the following case:

Case I

Smith knows that Hindley Earnshaw was Hareton's father, that Mrs Catherine Linton was the mother of Mrs Heathcliffe, and that Hindley and Mrs Linton were siblings. Smith wonders how Hareton and Mrs Heathcliffe are related. Smith thinks "Hareton's father was

> Hindley, and Mrs Linton was Hindley's sister, so Mrs Linton was Hareton's aunt; Mrs Heathcliffe is Mrs Linton's daughter, so Hareton and Mrs Heathcliffe are cousins." Consequently, Smith comes to believe that Hareton and Mrs Heathcliffe are cousins. Smith believes that Hareton and Mrs Heathcliffe are cousins because he inferred it. But Smith does not believe that they are cousins because they are cousins. Nevertheless, Smith knows that they are cousins.

Explanationism entails that this sort of case is impossible. But it seems to be possible.

Here is another case, inspired by a puzzle in (Smullyan 1982, 4):

> **Case II**
>
> Smith knows that Jones bought something for $7, sold it for $8, bought it back again for $9, and sold it again for $10. Smith wonders how much money Jones made trading the item. Smith thinks "-7 plus 8 is 1, 1 minus 9 is -8, -8 plus 10 is 2, so Jones made $2." Consequently, Smith comes to believe that Jones made $2. Smith believes that Jones made $2 because he inferred it. But Smith does not believe that Jones made $2 because Jones made $2. Nevertheless Smith knows that Jones made $2.

In order for Smith to *know* that Jones made $2, must the fact that Jones made $2 explain why Smith believes it? I think not.

§ 4.2.2 Bogardus & Perrin's Account of Deductive Knowledge

In this section I discuss Bogardus & Perrin's account of deductive knowledge and argue that one should sooner accept my counterexamples to the necessity of the condition than accept Bogardus & Perrin's account of deductive knowledge.

Bogardus & Perrin think that all possible cases in which someone comes to believe something because they inferred it, in which the inferring was a deducing, and in which all of the facts from which the person deduced it were known by them at the time of the deduction are cases in which the person believes the conclusion because they "saw" it (2022, 190). Call this

the "seeing" account of deductive knowledge. What is this "seeing"? I think it best to quote Bogardus & Perrin at length:

> Deductive arguments are a means by which we come to metaphorically "see" the truth of some proposition.…It's like an ordinary case of non-metaphorical vision. Vision is fallible, since it might be a hallucination or illusion. But so long as your visual system is functioning well, it puts you in a position to see objects before you, and if one such is a chair, you believe there's a chair there *because* there's a chair there. The fact that there's a chair there is a crucial part of the explanation of why you believe there is. Similarly, when you see the premises are true and the inference valid, deduction helps you see the conclusion. And that's because, in a sound argument, the conclusion is already "there," so to speak, in the premises. When one appreciates a sound argument, one is not merely seeing the truth of the premises and the validity of the inference; one also sees the truth of the conclusion thereby. In that case, you believe the conclusion because it's true. This is fallible, when the component beliefs or premises might be false or otherwise unknown. But, when they're true and you know it, then deduction positions you to see the truth of the conclusion (Bogardus & Perrin 2022, 190).

Bogardus & Perrin would argue that if Case I and Case II were possible there would be a 'seeing' in each case and in Case I Smith would believe that Hareton and Mrs Heathcliffe are cousins *because* they are cousins and in Case II Smith would believe that Jones made $2 *because* Jones made $2. They would argue that the stipulations to the contrary (e.g. "Smith does not believe that Jones made $2 because Jones made $2") make the cases impossible.

In the interests of brevity, I shall consider only the argument they would offer which pertains to Case II. What I am interested in in particular is the argument they would offer for the conclusion that if Case II were possible then Smith would believe that Jones made $2 because Jones made $2, i.e. I am interested in η:

η

1 Case II is possible
[Assumption]

2 Smith believes that Jones made $2 because he "saw" that Jones made $2.
[1, the "seeing" account of deductive knowledge]

3 Smith "saw" that Jones made $2 because Jones made $2.
[2, the nature of 'seeing'(?)]

4 Smith believes that Jones made $2 because Jones made $2.
[2 & 3]

If Case II were possible, then Smith would believe that Jones made $2 because Jones made $2.
[1–4]

I shall argue that if "seeing" is knowing/understanding, then Premise 2 is false. Then I shall argue that if 'seeing' is thinking/judging/concluding, then Premise 3 is false. Then I shall argue that if the 'seeing' is something other than Smith's understanding or Smith's concluding, we should sooner believe that Case II is a counterexample to the necessity of the condition than believe in the existence of 'seeing's; and if there are no 'seeing's, then Premise 2 and Premise 3 are both false. Hence, I conclude that η is unsound and Case II is a counterexample to the necessity of the condition.

First, what if 'seeing' is knowing/understanding? I assume that Smith's understanding that Jones made $2 just is his knowing that Jones made $2. I accept that Smith knows that Jones made $2, hence I accept that Smith understands that Jones made $2. So, in that familiar sense of "see" in which it means what "understand"/"know" means, Smith does see that Jones made $2. But Smith doesn't believe that Jones made $2 because he knows/understands that Jones made $2; nor does Smith believe that Jones made $2 because he knew/understood that Jones made $2. Smith's knowing/understanding that Jones made $2 is neither explanatorily nor temporally prior to Smith's believing that Jones made $2. So, if "seeing" is knowing/understanding Premise 2 is false.

Now suppose that Smith's "seeing" that Jones made $2 is his thinking "so Jones made $2"/"Jones made $2." Smith's concluding/judging/thinking "so Jones made $2"/"Jones made $2" is not analogous to a perceptual act in the way in which Bogardus & Perrin need it to be: the fact that Jones made $2 does not explain why Smith thought "so Jones made $2." It is just

as implausible that the fact that Jones made $2 explains why Smith thought "so Jones made $2"/"Jones made $2" as it is that the fact that Jones made $2 explains why Smith believes that Jones made $2. The 'seeing' was appealed to in order to account for how, if at all, the belief could be explained by the fact believed: something else would be needed in order to explain how, if at all, the judgment could be explained by the fact believed. A regress looms. So, if 'seeing' is concluding/judging/thinking, Premise 3 is false.

Could the 'seeing' be anything other than the understanding or the concluding? If it were anything else, I don't see why we should believe that there is such a thing. It seems to me that we don't need to appeal to the alleged fact that Smith "saw" that Jones made $2 in order to explain why Smith believes that Jones made $2: he believes it because he inferred it and because he believed the premises from which he inferred it; and no 'seeing's are needed to explain these facts. Explanationists must demonstrate the explanatory necessity of "seeing"s if we are to believe they exist. As of right now, they are hypothetical entities and the only reason for believing in them is that they would solve a problem for Explanationism. I therefore think that one should sooner accept Case II to be a counterexample to the necessity of the condition than accept that there are 'seeing's which aren't understandings or concludings.

So, either Premise 2 is false because 'seeing' is understanding, or Premise 3 is false because 'seeing' is concluding, or Premise 2 and Premise 3 are false because there are no 'seeing's. Hence one should sooner accept Case II to be a counterexample to the necessity of the condition than accept Bogardus & Perrin's 'seeing' account of deductive knowledge. The same is true of Case I.

§ 5. Conclusion

In this paper I have discussed some consequences of Explanationism. In § 3 I questioned the sufficiency of the condition. In § 4.1 I argued that there are skeptical consequences of the necessity of the condition. In § 4.2 I gave two counterexamples to the necessity of the condition which both concern deductive knowledge.

Acknowledgements

I thank a reviewer for having been very generous with their time and for having read the paper carefully. Their suggestions and comments led to my making substantial improvements to the paper. I thank the editors for giving me a chance to improve the paper. I thank Maria Zanella for reading drafts of the paper, for inventing the specific example given in § 3.3.2, and for helpful conversations about knowledge.

References

Baron-Schmidt, Nathaniel. 2024. "Thing causation." *Noûs*, 58(4): 1050–72. https://doi.org/10.1111/nous.12494

Bogardus, Tomas. 2014. "Knowledge Under Threat." *Philosophy and Phenomenological Research*, LXXXVIII(2). 289–313. https://doi.org/10.1111/j.1933-1592.2011.00564.x

Bogardus, Tomas and Will Perrin. 2022. "Knowledge is Believing Something Because It's True." *Episteme*, 19. 178–96. https://doi.org/10.1017/epi.2020.18

Bogardus, Tomas and Will Perrin. 2023. "A Defense of Explanationism against Recent Objections." *Episteme*. 1–12. https://doi.org/10.1017/epi.2023.42

Boyce, Kenneth and Andrew Moon. 2023. "An Explanationist Defense of Proper Functionalism." In *Externalism About Knowledge*, edited by Luis R. G. Oliveira, 277–300. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198866749.003.0011

Carrier, L. S. 1976. "The Causal Theory of Knowledge." *Philosophia*, 6(2): 237–57. https://doi.org/10.1007/BF02379925

Colaço, David, Wesley Buckwalter, Stephen Stich, and Edouard Machery. 2014. "Epistemic Intuitions in Fake-Barn Thought Experiments." *Episteme*, 11(2): 199–212. https://doi.org/10.1017/epi.2014.7

Gendler, Tamar Szabó and John Hawthorne. 2005. "The Real Guide to Fake Barns: A Catalogue of Gifts for Your Epistemic Enemies." *Philosophical Studies*, 124: 331–52. https://doi.org/10.1007/s11098-005-7779-8

Gettier, Edmund L. 1963. "Is Justified True Belief Knowledge?" *Analysis*, 23(6): 121–23. https://doi.org/10.1093/analys/23.6.121

Goldman, Alvin. 2009. "Williamson on Knowledge and Evidence." In *Williamson on Knowledge*, edited by Patrick Greenough & Duncan Pritchard, 73–91. Oxford: Oxford University Press. https://doi.org/10.1093/aprof:oso/9780199287512.003.0006

Goldman, Alvin I. 1967. "A Causal Theory of Knowing." *The Journal of Philosophy*, 64(12): 357–72. https://doi.org/10.2307/2024268

Goldman, Alvin I. 1976. "Discrimination and Perceptual Knowledge." *The Journal of Philosophy*, 73(20): 771–91. https://doi.org/10.2307/2025679

Lehrer, Keith and Thomas Paxson, Jr. 1969. "Undefeated Justified True Belief." *The Journal of Philosophy*, 66(8): 225–37. https://doi.org/10.2307/2024435

Mortini, Dario. 2022. "The Explanationist and the Modalist." *Episteme*. 1–16. https://doi.org/10.1017/epi.2021.57

Piccinini, Gualtiero. 2022. "Knowledge as Factually Grounded Belief." *American Philosophical Quarterly*, 59(4): 403–17. https://doi.org/10.5406/21521123.59.4.06

Plantinga, Alvin. 1988. "Positive Epistemic Status and Proper Function." *Philosophical Perspectives*, 2: 1–50. https://doi.org/10.2307/2214067

Plantinga, Alvin. 1993. *Warrant: The Current Debate*. Oxford: Oxford University Press. https://doi.org/10.1093/0195078624.001.0001

Schellenberg, Susanna. 2018. *The Unity of Perception: Content, Consciousness, Evidence*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198827702.001.0001

Smullyan, Raymond. 1982. *The Lady or the Tiger?* Oxford: Oxford University Press. [Reissue from 2000] https://archive.org/details/ladyortiger00smul/page/16/mode/1up [Alternative Edition]

Sosa, Ernest. 1999. "How to Defeat Opposition to Moore." *Philosophical Perspectives*, 13: 141–53. https://doi.org/10.1111/0029-4624.33.s13.7

Starmans, Christina and Ori Friedman. 2012. "The Folk Conception of Knowledge." *Cognition*, 124: 272–83. https://doi.org/10.1016/j.cognition.2012.05.017

Starmans, Christina and Ori Friedman. 2020. "Expert or Esoteric? Philosophers Attribute Knowledge Differently Than All Other Academics." *Cognitive Science*, 44: 1–31. https://doi.org/10.1111/cogs.12850

Strevens, Michael. 2011. *Depth: An Account of Scientific Explanation*. Cambridge, Massachusetts: Harvard University Press.

Turri, John. 2016. "Knowledge and Assertion in 'Gettier' Cases." *Philosophical Psychology*, 29(5): 759–75. https://doi.org/10.1080/09515089.2016.1154140

Organon F 32 (2) 2025: 224–259 ISSN 2585-7150 (online)
https://doi.org/10.31577/orgf.2025.32205

RESEARCH ARTICLE

On Two Underappreciated Motifs of Quine's Naturalism; Or, Quine on Reality and Naturalistic Philosophy

Michal Hubálek*

Received: 14 August 2024 / Revised: 21 February 2025 / Accepted: 5 March 2025

Abstract: This essay deals with two often underappreciated Quinean naturalistic motifs. The first motif concerns the metaphysical status of reality under naturalism, by examining what can be viewed as Quine's attempt to dissolve realism and idealism into each other; the second motif concerns the metaphilosophical dimension of naturalized epistemology. I aim to demonstrate that Quine's way of approaching these two motifs turns out to be not only complementary, but also highly indicative of his overall naturalistic outlooks as well as professional preferences. As a result, this essay re-assesses Quine's influence on the development of naturalism and naturalistically conceived philosophy from the second half of the 20th century onwards.

Keywords: Quine; realism; naturalism; naturalized epistemology; naturalistic philosophy.

Introduction

Naturalism is nowadays without a doubt understood in many different ways (see, e.g., Bryant 2020; Rosenberg 1996; cf., Danto 1967). Part of the

* University of Hradec Králové
https://orcid.org/0000-0002-9089-3556
Department of Philosophy and Social Sciences, University of Hradec Králové, Rokitanského 62/26, 500 03 Hradec Králové 3, Czech Republic
hubalek.michal.42@gmail.com

reason for this may be the fact that the subscription to at least a very broad naturalism is a standard for present-day philosophers. As David Papineau pointedly states: "For better or worse, 'naturalism' is widely viewed as a positive term in philosophical circles – only a minority of philosophers nowadays are happy to announce themselves as 'non-naturalists,'" (Papineau 2021). So, in recent years, we have witnessed several instructive attempts to (re)define naturalism anew (I choose Price 2011 and De Caro & Macarthur 2004). Many of these attempts go, whether directly or indirectly, in the footspets of the two well-known naturalists writing in the second half of the 20th century, W. V. O. Quine and Wilfrid Sellars (see, e.g., Peregrin 2023; Roth 2020). The aim of this essay is to contribute to contemporary debates about naturalism by dealing with two often underappreciated motifs of Quine's naturalism.

In the first part of the essay, I focus on the first motif which concerns the metaphysical status of "reality" under naturalism, by examining what can be viewed as Quine's attempt to dissolve realism and idealism into each other. In the second part, I move on to parse the second motif which concerns the meta-philosophical dimension of naturalized epistemology.

Moreover, as will gradually become evident, I present Quine's naturalistic legacy in a manner that makes Quine much more complementary to thinkers such as John Dewey, Ludwig Wittgenstein, Wilfrid Sellars, Stephen Jay Gould, or Richard Rorty than usually acknowledged in print. Put another way, although there are certainly many differences between Quine and the thinkers just mentioned, the similarities highlighted in this essay, however local they might seem, challenge the widespread view that these thinkers radically broke away from Quine's very own naturalistic framework.

1. Dissolving Realism and Idealism into Each Other

> *Natural selection the great solvent. Dissolved final cause into efficient cause. Dissolves realism and idealism into each other. Dissolves disparateness of neural input into a like output. Kinds are projected for terms, and terms are vague, applying in degrees.*

> *Relative similarity is what is objective; not natural kinds. Biological species are atypically clean-cut. Similarity is projected on the world from our associations.*
>
> –W. V. O. Quine[1]

1.1 A Realist Quine

A lot has been written on Quine and real*ism* and no wonder (see, e.g., Jaksland 2024; Pils 2020; Keskinen 2012; Dreben 1992; Rorty 1976).[2] Let me start right in the middle and introduce the basics of Quine's position on this matter. Quine's naturalistic, intra-theoretic position holds that the "objects" laymen as well as scientists talk about (e.g., atoms, genes, stones, bodies, and numbers) are to be regarded as linguistic posits. Qua posits, objects of our discourse are then seen as conceptual tools we develop and utilize to accommodate and make sense of our experiences (of our causal interactions with the world). Then, for both laymen as well as scientists, any idea of "reality" must be worked out in a theory-internal way.

We inherit and invent concepts and theories about the world which commit us to a certain ontology; they commit us to positing certain kinds of objects. For us humans, the existence of specific objects with their specific characteristics is thus always theory-dependent. To ask what objects

---

[1] (Quine 2008, 179–80)

[2] I purposely refrain from providing the reader with a concrete delimitation of realism and anti-realism here to effectively highlight the fact that Quine's own position is quite idiosyncratic. A very minimalistic delimitation should do; *The Oxford Companion to Philosophy* instructively states: "[Realism and anti-realism are] primarily directions, not positions. To assert that something is somehow mind-independent is to move in the realist direction; to deny it is to move in the opposite direction. No sane position is reached at either extreme. Not everything is in every way independent of minds; if there were no minds, there would be no pain. Not everything depends in every way on minds; if I forget that Halley's comet exists, it does not cease to exist. Many philosophical questions have the general form: Is such-and-such mind-independent in so-and-so way? Given specifications of such-and-such and so-and-so, one may call someone who answers 'Yes' a realist. Since different philosophers take different specifications for granted, the word 'realism' is used in a bewildering variety of senses…" (Honderich 1995, 746–47).

there are independently of any concepts and theories is hopeless, according to Quine. Without theories there are no things-in-themselves, but only blooming, buzzing confusion (Quine 1995a, 225). Also, due to Quine's well-known commitment to holism, the identity conditions for particular objects are dependent on a bundle of theories taken as a whole; they depend on *science as a whole*. Quine often also speaks of our "web of belief" or "total theory of the world" (Quine 1961a; 1961b; cf., Quine & Ullian 1978 and Quine 1987, 108ff). Note that this all applies to "physical objects" as well. In this sense, physical objects are also science-dependent:

> Physical objects are conceptually imported into the situation as convenient intermediaries – *not by definition in terms of experience*, but simply as irreducible posits comparable, epistemologically, to the gods of Homer. Let me interject that for my part I do, qua lay physicist, believe in physical objects and not in Homer's gods; and I consider it a scientific error to believe otherwise. *But in point of epistemological footing the physical objects and the gods differ only in degree and not in kind. Both sorts of entities enter our conception only as cultural posits.* The myth of physical objects is epistemologically superior to most in that it has proved more efficacious than other myths as a device for working a manageable structure into the flux of experience.…
> Total science, mathematical and natural and human, is…extremely underdetermined by experience. *The edge of the system must be kept squared with experience; the rest, with all its elaborate myths or fictions, has as its objective the simplicity of laws.* (Quine 1961a, 44, 45, emphasis mine)

For many thinkers, these postulations have anti-realist or even idealist undertones. Nevertheless, Quine still wants to be an empiricist; he holds that for an experience to be *evidence of/for* something there must be some inter-subjective, observable/behavioral criteria at work. So, the objects, to whose existence we are committed, are in fact dependent not only on our theories but also on "the world" (given causal contexts) (see Quine 1969a; 1992, 5ff; cf., 1987, 159–161). Quine of course never denies this, ipso facto, he never denies that *the world exists independently of us*. The only difficulty here is that, according to him, our human knowledge of this world is co-determined

by our conceptual choices and the facts of the matter at the same time, with no unambiguous way of separating these two elements. This is a result of dismantling the fundamental distinction between analytic and synthetic beliefs/truths (cf., Roth 2020, 115, 169, note 2; and Danto 1967).

One of the early commentators dealing with Quine's holistic view of language and scientific knowledge, Albert Hofstadter, who moreover pointedly labels Quine's position as "holistic pragmatism," contemplates that, in Quine's theoretical system, there is no room for a philosophically significant distinction between scientific invention and discovery. And although Hofstadter in the end to a large degree diverges from Quine on this point, he instructively summarizes Quine's outlook as follows:

> The consequence for the invention-discovery distinction is that while science depends upon both human and non-human factors, one can not distinguish the two factors within science.…Both operate *in* a kind of functional relation to each other and *on* the whole of the language of knowledge. There are no empirical tests or conventional decisions regarding statements individually. Every test and every decision reverberates through the whole, affects and is affected by the whole, so that in every empirical test the whole of our knowledge is tested and in every decision the whole fabric is decided on. *Or perhaps better, the distinction between test and decision also breaks down*" (Hofstadter 1954, 401; cf., 412ff).

One could say for now that, for Quine, our world, the world we live and theorize in, is both invented and discovered; made and found.

However, especially due to Quine's insistence on the observable/behavioral notion of evidence, is not Quine then committed to a position taking, e.g., observable "items" (however holistically conceived) as in some sense primary, or even as the ones that are *really real*? The answer to this question is negative, as Quine explicitly subscribes to a view that all objects are theoretical objects. If there is a difference between positing the existence of moles and molecules, it is a difference in degree and not in kind (e.g., Quine 1961c, 18; Quine 1969a, 87). Again, without our sentences (or generally conceptual frameworks), there would not be any objects or facts for us to talk about (see, e.g., Quine 1981a, 20). But, interestingly enough, this does

not discourage Quine from declaring, in the next breath, that he endorses "robust realism," i.e., "unswerving belief in external things – people, nerve endings, sticks, stones." And then continues by stating: "I believe also, if less firmly, in atoms and electrons and in classes," (Quine 1981b, 21). But the belief in physical objects is firmer, among other things, only insofar as the words for them are acquired ontogenetically *earlier* than for abstract objects when one acquires a first language, so physical objects are "more deeply rooted in our formative past." In this particular regard, the vocabulary of physical objects is considered primitive, as it is directly tied to our interactions with the causal realm (Quine 2013, 215).

1.2 Naturalistic Reconfiguration of "Reality"

With respect to the traditional disputes about the philosophical forms of realism and anti-realism, Quine finds himself in an outlandish situation insomuch as he prima facie saves the most fundamental beliefs of both of these isms; Quine does not deny that the world itself (the causal realm) exists independently of us, and simultaneously, grants us significant creative freedom in imposing our conceptual frameworks on this world.

Nonetheless, insomuch as Quine is resolute in denying the theory-*independent* existence of objects for us human cognizers and agents, I contend that he is much closer to anti-realism than to realism, at least to any naïve forms of realism; for Quine, the identity conditions and/or truth values for categorizing and describing objects are theory-*dependent*. Also, I think that the vast majority of realists would want to hear more, for example, that the world possesses its own unifying and/or knowable structure as well as rules/principles. In other words: its own integrity (cf., Wright 1993).

I thus want to proceed by defending Quine, even if somewhat indirectly, against the accusation that his holism forces him to embrace a form of idealism. Hofstadter, for example, adverts Quine's "holistic pragmatism" as "an empiricist equivalent of the Idealistic Absolute" (Hofstadter 1954: 416). Doing this will, I hope, help to establish what kind of anti-realist Quine actually is. Because indeed, according to Quine, the *reality* we recognize and live in is co-created by us, because "all ascriptions of *reality* must come…from within one's theory of the world; it is incoherent otherwise," (Quine 1981a, 21). Quine a bit later sums up his views in this way:

> *The objectivity of our knowledge of the external world remains rooted in our contact with the external world, hence in our neural intake and the observation sentences that respond to it.* We begin with the monolithic sentence, not the term. A lesson of proxy functions is that our ontology, like grammar, is part of our own conceptual contribution to our theory of the world. *Man proposes; the world disposes*, but only by holophrastic yes-or-no verdicts on the observation sentences that embody man's predictions (Quine 1992, 36, emphasis mine).

It is obvious, and fully comprehensible from a naturalistic vantage point, that Quine's "solution" (I use quotation marks, as Quine himself never addresses this issue directly) to the problem of "reality" is motivated by understanding our human position in the world; *our human epistemic condition*. We are always *in* the world and all our beliefs about this world are always determined by our cognitive faculties and vocabularies interacting with the world. After all, that is what our current science (numerous branches of biology and anthropology) tells us. We are animals/beings sharing the world with other animals/beings. We "antecedently acknowledge" the external world in everything we do, hence also when inquiring into and describing the world (Quine 1992, 19).

One must always take into account that Quine thought of himself as a Darwinian (see, e.g., Quine 1996). However, in this particular regard, Quine is also very close to the pragmatism/naturalism of John Dewey who, inter alia, forcefully advanced an anti-representational view of our knowledge of the world (cf., Quine 1969d, 27ff). Dewey was convinced that many thinkers, and philosophers especially (see Dewey 1911), make a pragmatically impotent and theoretically unnecessary question from the relationship between our cognition and the world. Louis Menand fruitfully summarizes Dewey on this point as follows:

> The pragmatist response to this question is to point out that nobody has ever made a problem about the relation between, for example, the hand and the world. The function of the hand is to help the organism cope with the environment; in situations in which a hand doesn't work, we try something else, such as a foot,

> or a fishhook, or an editorial.…*Dewey thought that ideas and beliefs are the same as hands: instruments for coping. An idea has no greater metaphysical stature than, say, a fork.*…Dewey's point was that "mind" and "reality," like "stimulus" and "response," name nonexistent entities: they are abstractions from a single, indivisible process. It therefore makes as little sense to talk about a "split" that needs to be overcome between the mind and the world as it does to talk about a "split" between the hand and the environment, or the fork and the soup (Menand 2002, 360–361, emphasis mine).

This is a broadly naturalistic approach, since it openly draws on the Darwinian insight that we are ultimately organisms *coping* with our environments, and not necessarily *copying* its metaphysical structures (cf., Rorty 2021; 1998). With this starting point, it is very hard to conceive an argument claiming that we might not be in touch with the (external/real) world. The only possibility seems to be claiming that we are not in touch with the world *as it is in itself*. But I do not think that there actually is a conclusive counter-argument to this Kantian move, i.e., an argument that makes a pragmatic difference to our everyday life and/or various forms of theorizing. It will always be possible to insist on the possibility that there is a "really real" world we cannot reach for some reason or other.

In view of this, I argue that one cannot satisfactorily answer the question "what is the metaphysical status of reality in Quine's theoretical system?," unless one takes seriously his naturalistic viewpoints. That is to say, our relationship to the world is, in principle, empirically explainable, meaning causally and/or historically (genealogically). In contemporary idiom, one could say that the inquiries into our phylogeny, ontogeny, and cultural history offer the most fundamental understanding of the human condition in the world.[3]

[3] However, this is not to say that these inquiries produce some "objectively" fundamental starting points. They are obviously also scientific theories. In other words, e.g., while developing specific evolutionary explanations, scientists make many conceptual choices and commitments, at minimum with respect to "science" and "evolution." Also, as the survey study of Tuomisto et al. (2018) tentatively shows, the inquiry into the origin of exclusively human traits across the scientific fields such

Just like Darwin, Quine does not strive to identify an essence of human beings/cognizers; his strategy is ethological and ecological; he is interested in behavioral patterns and principles typical for us as a biological and social species, which are the result of our modus vivendi (of our living conditions). This directly implies that we, as humans, always approach the world *from within our theories about it*. There is neither a profound nor fundamental level of the world. We cannot go beyond a posited world to understand *who we are* and *what world* we actually inhabit. This is a conceptual leap most of the naïve realists will simply not accept.[4] This leap, moreover, undermines any philosophical, and I would also say any pragmatic, substance of the distinction between *the world in itself* and the *world for us*.

For Quine himself, the most valuable naturalistic insight into "the nature" of the world comes from the study of our acquisition of first language and of our language use in general, since any genuine observation is a linguistic matter (it presupposes inductive agreement on our basic categories). Inter alia for this reason, our common sense, science, and rationality itself comprise one cognitive and epistemic continuum. As is well known, Quine

as paleoanthropology, paleontology, ecology, evolution, and human biology is very diverse, meaning that it is possible to identify popular hypotheses, but there is no universal consensus. Joseph Heath and Catherin Rioux in the same vein reflect on the current state of affairs in evolutionary ethics and write: *A common mistake made by philosophers working under the banner of "evolutionary ethics" has been to suppose that the evolutionary science on this question is settled* – that we have an answer to the question, not just how human morality is possible, but even how it evolved. *If this were correct, then the only task remaining for the philosopher would be to draw out the normative implications of this body of science. Unfortunately, there is no such scientific consensus. Indeed, the evolution of human ultrasociality – whether it be altruistic or cooperative – is one of the most important unanswered questions in the life sciences*," (Heath and Rioux 2018, 1–2, emphasis mine). Of course, it is also necessary to ask whether the consensus on these matters is possible/desirable, but I leave aside that here.

[4] See, for example, the recent attempt to endorse naïve realism by Mitrović (2022). Typically, as treated again in *The Oxford Companion to Philosophy*, naïve realism amounts to "a theory […] that holds that our ordinary perception of physical objects is direct, unmediated by awareness of subjective entities, and that, in normal perceptual conditions, these objects have the properties they appear to have…," (Honderich 1995, 602).

is not interested in any ahistorical demarcation criterion of science (see especially Quine 1995a; 1957).

That is the proper context for understanding Quine's much-quoted passages on ontology from his notorious essay "On What There Is." Consider now these samples:

> *To be assumed as an entity is, purely and simply, to be reckoned as the value of a variable.* In terms of the categories of traditional grammar, this amounts roughly to saying that *to be is to be in the range of reference of a pronoun.* Pronouns are the basic media of reference; nouns might better have been named propronouns (Quine 1961c, 13, emphasis mine).

> We can very easily involve ourselves in ontological commitments by saying, for example, that *there is something* (bound variable) which red houses and sunsets have in common; or that *there is something* which is a prime number larger than a million. But, this is, essentially, the *only* way we can involve ourselves in ontological commitments: by our use of bound variables (ibid., 12).

> Viewed from within the phenomenalistic conceptual scheme, the ontologies of physical objects and mathematical objects are myths. *The quality of myth, however, is relative; relative, in this case, to the epistemological point of view. This point of view is one among various, corresponding to one among our various interests and purposes* (ibid., 19).

However, it is easy to underestimate, or just forget, that Quine's theoretical interests and interpretations are constantly driven by pragmatic hence primarily epistemic concerns. Ontology follows epistemology, not vice versa. Or, as one of the two epigraphs for *Word and Object* (2013) has it; "ontology recapitulates philology."[5] Quine memorably opens his essay "Ontological Relativity" by this announcement: "With Dewey I hold that knowledge, mind, and meaning are part of the same world that they have to do with, and that they are to be studied in the same empirical spirit that animates natural science" (Quine 1969d, 26). This announcement, however, can be read in several ways, each of which is relative to a particular understanding

[5] The author of the epigraph is an American biologist James Grier Miller.

of the key concepts ("the world," "empirical," and "natural science") as well as to our interests and purposes. Put another way, although it is not obvious on first reading, or on contextless reading, it is a conceptually hence ontologically plastic announcement.

The ontological relativity, as Quine construes it, arises from the appreciation that

> there be no ultimate sense in which [the] universe can have been specified.… *Ontology is indeed doubly relative*. Specifying the universe of a theory makes sense only relative to some background theory, and only relative to some choice of a manual of translation of the one theory into the other.…We cannot know what something is without knowing how it is marked off from other things. *Identity is thus of a piece with ontology*," sums up Quine (1969c, 50, 54–55, emphasis mine).

Also, nothing prevents us in these settings from utilizing the Wittgensteinian conception of "language games," as Quine himself has done occasionally since the 1990s (see, e.g., Quine 1992, 20).

Now, to rebut the reading of Quine as being an idealist of a solipsistic kind, one must ask how the causal realm (the world) constrains our linguistic practices of posit*ing* objects. One also must agree with Davidson that: "It is reasonable to call something a posit if it can be contrasted with something that is not. [For Quine] the something that is not is sensory experience – at least that is the idea," (Davidson 1973–1974, 16). But since Quine denies that there is a possibility to unambiguously detect the empirical content of individual fact-stating terms and sentences, one must reformulate the question.

Let me try. For starters, it would be more useful to ask, "what makes some posited objects *better* than others and/or what makes them *genuinely scientific*?" (meaning simply what distinguishes them from our everyday/common-sense positing).

The straightforward (but only partial) answer would be, of course, that the answer to this question will always remain relative to our interests and purposes, but Quine in fact offers more than that. And, in this particular sense, Quine's dealing with the problem of positing is a special instance of (scientific) hypothesizing (cf., Quine 1995b, 49–50). Quine appreciated

very soon that abductive inferences, or what is sometimes called "the inference to the best explanation," is a very useful and in fact indispensable tool for our hypothesis making.

> Hypothesis, where successful, is a two-way street, extending back to explain the past and forward to predict the future. What we try to do in framing hypotheses is to explain some otherwise unexplained happenings by inventing a plausible story, a plausible description or history of relevant portions of the world

writes cogently Quine and Ullian (1978, 66). An instructive example at several levels is Quine's dealing with the concepts of "meaning" and "mental state" (believing, wishing, intending, etc.). Nowadays, it is hardly controversial that Quine refuses mind-body dualisms of Descartes' type and subscribes to "anomalous monism" (see Davidson 2002): "The point of anomalous monism is just that our mentalistic predicate imposes on bodily states and events a grouping that cannot be defined in the special vocabulary of physiology. Each of those individual states and events is physiologically describable, we presume, given all pertinent information," explains Quine (1995b, 88; cf., Quine 1992, 71ff).[6] However, *as a naturalist*, Quine cannot simply dogmatically start with this kind of physicalist ontology. Such an ontology must be a result of empirical/scientific considerations in order to be a naturalistic ontology.

So, the question here for us ultimately is: "why does Quine not think that meanings and mental states should be considered as (scientific) posits?" Roth offers a lucid summary of Quine's take on this matter in the form of an argument:

Pr. 1: For posits to have claim to reality, they must be justified as part of an explanatory theory within a naturalized epistemology.

[6] Quine is very consistent on this point, consider this telling passage of his: "Descartes' dualism between mind and body is called metaphysics, but it could as well be reckoned as science, however false. He even had a causal theory of the interaction of mind and body through the pineal gland. *If I saw indirect explanatory benefit in positing sensibilia, possibilia, spirits, a Creator, I would joyfully accord them scientific status too, on a par with such avowedly scientific posits as quarks and black holes*," (Quine 1995a, 252, emphasis mine).

Pr. 2: In order to be a justified part of an explanatory theory, the posits must either be necessary for (constitutive of) stimulations being evidence (as in quality spaces), or must provide an observable mark marking them as the things they are, i.e., there must be an objective basis for ascertaining that some behavior has been rightly or wrongly categorized a behavior of a certain type,

Pr. 3: But mental states and concepts have yet to be shown to be necessary for or even robustly explanatory of observed stabilities in behaviors; in addition, nothing marks a behavior as objectively categorized correctly or incorrectly,

C: Meanings and mental states have no claim to reality (because no claim to any genuine explanatory function). Whatever has no place in our best scientific scheme of explanations is not real. Whatever is not real does not belong to the realm of facts, i.e., constitute a fact of the matter. (Roth 2003, 275)

Premise number two has principal significance for understanding Quine's concept of posit. "Posits can be constitutive of the possibility of stimulations being evidence, or they can be kinds made objective by accepted theoretical explanations," (ibid.). The only objective, better said intersubjective, criteria for ascribing mental states to people are observable, i.e., physicalist/behavioral, criteria.[7] Note that this is merely an alternative way of acknowledging the "reality" of sensory experience, as Davidson alludes to in the quote above.

[7] "Quine charts what falls by the way in an advancing scientific picture of what there is. His view of epistemology as science self-applied, and his corresponding conclusion about the explanatory utility of meanings, has interesting parallels here to Kantian themes. *For at least one important link that connects a tradition that runs from Kant to Quine studies how human minds come to constitute a shared and mutually intelligible world. Kant, in this regard, assigns the inquiring mind a strongly constitutive role. But how concepts and percepts come together to form understanding remains an unsolved problem, a problem which finds its modern incarnation in questions of finding rules that guide behavior. Quine despairs of the task of using philosophical analysis to find such rules. He proposes, instead, to turn that job over to empirical psychology*" (ibid., 278, emphasis mine).

What we then actually do when speaking about mental states is *imposing mentalistic predicates on bodies*. Quine's argument is thus not reductionist, nor *primarily* ontological. It is a pragmatic and explanatory argument. So, according to Quine, to say that "ontology recapitulates philology" also means that *ontology recapitulates epistemology*.

To put it crudely, Quine argues that when explaining or interpreting human behavior, one does not need to assume "meanings" and "mental states" as ontologically distinct entities existing somewhere in the causal world (or in our heads, for example). However, he does not directly reject the idea that mentalistic or intentional vocabulary can sometimes serve as a satisfactory means for explaining and interpreting particular cases.

That is the key pragmatic difference here; to work with mental states (ascribing them to people) might be useful for interpreting human action, but positing a sui generis ontology of the mental does not, in itself, add to their explanatory force (unless, of course, positing an autonomous realm of the mental proves to be explanatorily beneficial, too). Recall that, Quine ultimately holds that all vocabularies function "merely" as conceptual tools for navigating our experiences of the world. This stance certainly differentiates Quine from other influential naturalists, e.g., from Wilfrid Sellars, when it comes to its ontological implications. However, how much difference this stance makes for our conception of empirical inquiry as such remains an open and often-discussed question (to learn more, one can start with Rosenberg 2007; Borradori 2008; and Roth 2023).

For the purposes of this chapter, the key point is that Quine's epistemology-first stance intricately aligns ontology with empirical and pragmatic concerns.

1.3 Fusing Idealism with Instrumentalism

At this point, another way of resuscitating the realist element of Quine's ontology reappears. Quine could be seen as a *scientific* realist. It could be argued that, because science is ultimately very successful technologically, it must provide us with at least some truths about the real nature of the world around us.

However, in Quine's intra-theoretic/intra-linguistic system, such a realism loses its metaphysical bite and is still only a form of instrumentalism,

meaning a *methodological* position emphasizing the usefulness of posits/concepts qua tools. Be that as it may, Quine is very far from admitting the meaningfulness of the Platonic assumption of seeing our theories as "carving nature at its joints." Strictly speaking, for Quine, one can only "know" (intelligibly state) that there is nature (independent causal realm); whether nature has *its own* joints cannot be known unambiguously. It is simply hard to imagine what it would mean to know that *in pragmatic terms*. And to proclaim our "best" scientific posits as really real is at least problematic, due to all kinds of semantic indeterminacies and evidential underdeterminations Quine famously introduces (cf., e.g., Quine 1975a, 302ff). If the history of science teaches us anything, it is the fact that from both synchronic and diachronic perspectives, there is more than one way of conceptually carving the world.

To summarize: according to Quine, our view of the world with its particular objects is always interwoven with our theoretical and other pragmatic choices.

> *To call a posit a posit is not to patronize it.…Everything to which we concede existence is a posit from the standpoint of a description of the theory-building process and simultaneously real from the standpoint of the theory that is being built. Nor let us look down on the standpoint of the theory as make-believe; for we can never do better than occupy the standpoint of some theory or other, the best we can muster at the time.* What reality is like is the business of scientists, in the broadest sense, painstakingly to surmise; and what there is, what is real, is part of that question. The question how we know what there is is simply part of the question…. *The last arbiter is so-called scientific method, however amorphous* (Quine 2013, 20–21, emphasis mine).

On my reading then, Quine performs a flight *from realism,* but it is not a flight *from reality* (hence nor *to any kind of idealism*). The metaphysical nature of our world and its objects are indeed "only" hypothesized; but the hypothesizing is our practical doing, it is hypothesizing via and against our interactions with the world/the causal contexts.

So, despite Davidson's intermittent suspicions, Quine would quite happily condone that:

> In giving up dependence on the concept of an uninterpreted reality, something outside all schemes and science, we do not relinquish the notion of objective truth – quite the contrary.…*In giving up the dualism of scheme and world, we do not give up the world, but re-establish unmediated touch with the familiar objects whose antics make our sentences and opinions true or false* (Davidson 1973–1974, 20, emphasis mine; cf., Quine 1984).

True, Quine's holistic notion of evidence implies, inter alia, that no single experience or scientific experiment will ever resolve an ontological dispute (cf., Quine 2013, 254). But why would one even expect that? It is certainly not licensed, for example, by our everyday experiences. Quite the opposite suggests itself. Consider the following quote from Roth, who briefly but pointedly explicates, in the Quinean/naturalistic attitude, what depending on theoretical and pragmatic choices means for our lives and our (scientific) understanding of the world:

> Background beliefs regarding social status or religious affiliation might influence which individual beliefs count or how they count. In addition, which beliefs might be open to revision will be determined by perceptions regarding how those beliefs connect to religious or political views deemed important. *Consideration such as these makes the "unit of empirical significance" culture-size* (Roth 2020, 125).

Of course, Quine's, in a sense instrumental, belief in/accepting of physicalism is also of this parochial kind. Physicalism, in Quine's view, keeps our science going; it provides science with many sometimes verifiable and sometimes falsifiable *hypotheses* on which we then gradually construct the rest of our web of belief/our total theory of the world.

Thus, although "we are now seeing ontology as a more utterly human option than we used to," the options are not infinite, precisely because we are regulated by the world and limited by our natural capabilities (Quine 1995a, 260). Naturalism in fact embraces these limitations. "Naturalism need not cast aspersions on irresponsible metaphysics, however deserved, much less on soft sciences or on the speculative reaches of the hard ones, except insofar as a firmer basis is claimed for the experimental method

itself" (ibid., 252, "experimental method" here simply stands for the inter-subjective/third-person perspective).

Therefore, contra Hofstadter's interpretation, nothing in Quine justifies claiming that "whatever view I want is a view of the Universe." A naturalistic corrective of this idealistic vision of the world could be: "There is the view of the universe which sees it as it is."[8] However, an obligatory note for those still under the spell of Kantian visions: here, "as it is" refers to "as it is for us, the beings of the universe," rather than "as it is in itself."

2. Naturalizing Philosophical Inquiry

> I see philosophy not as an a priori propaedeutic or groundwork for science, but as continuous with science. I see philosophy and science as in the same boat – a boat which, to revert to Neurath's figure as I so often do, we can rebuild only at sea while staying afloat in it. There is no external vantage point, no first philosophy. All scientific findings, all scientific conjectures that are at present plausible, are therefore in my view as welcome for use in philosophy as elsewhere.
>
> –W. V. O. Quine[9]

2.1 The Continuity of Philosophy and Science

Despite many misunderstandings, Quine's naturalistic approach to various forms of inquiry is not by any means hostile towards philosophy itself. Even philosophy is part and parcel of his naturalized epistemology. Philosophy only loses its special, grounding status towards science as Quine does not think that there is any purely philosophical method, meaning *epistemologically privileged philosophical perspective*, such as first-person perspective, introspective insight, etc. Methodologically-wise, *philosophy is continuous with science* (e.g., Quine 1969a; 1995a). As I also revealed in the previous section, for Quine, all forms of explication and explanation are science-dependent, or one could say result-driven; they must be formulated from

[8] I borrow these two characterizations from Peregrin (1999).

[9] (Quine 1969a, 126–27)

within a third-person perspective, otherwise the intersubjective, direct or indirect empirical checks, would simply not be available.

Philosophy thus also relies on the synchronic and diachronic developments of our web of belief; of our total theory of the world. So, Quine generally seems to oscillate between saying that philosophy and science are continuous (when he speaks in a disciplinary and methodological sense) and saying that philosophy is a part of science (when he means by science "our total theory of the world"). Consider in this light these two paragraphs:

> The naturalistic philosopher *begins his reasoning within the inherited world theory as a going concern.* He tentatively believes all of it, but believes also that some unidentified portions are wrong. *He tries to improve, clarify, and understand the system from within.* He is the busy sailor adrift on Neurath's boat (Quine 1981a, 72, emphasis mine).

> *Naturalistic philosophy is continuous with natural science. It undertakes to clarify, organize, and simplify the broadest and most basic concepts, and to analyze scientific method and evidence within the framework of science itself.* The boundary between naturalistic philosophy and the rest of science is just a vague matter of degree (Quine 1995a, 256–257, emphasis mine).

However, the pragmatic and intellectual purposes of philosophy as a distinct discipline are not specified in greater detail by Quine himself. As is clear from the quotes above and from one of the epigraphs I have chosen for this essay, Quine often only casually lists the general purposes of explication (clarifying and organizing concepts), methodological considerations, and aiming at understanding our web of belief/our total theory of the world as a whole. All the same, Quine can be read simply as emphasizing that philosophy can continue doing what it has always been doing as long as it respects the practices and results of other relevant disciplines and their theories, i.e., Quine can be read as not restricting the specific aims of philosophy in naturalism in advance of formulating given pragmatic or intellectual needs (cf., Churchland in Quine 2013, xiv).[10]

[10] Quine at one point also likens his idea of *philosophical explication* to Wittgenstein's vision of philosophy. He writes: "According to an influential doctrine of

On the other hand, I think that one must be very careful and always distinguish between the very consequences of Quine's naturalism for philosophy, and Quine's own philosophical preferences and personal tastes. From a certain perspective, Quine is a conservative thinker immersed in analyzing a very specific subset of scientific and philosophical problems, where the latter are separated from the former only by their higher degree of generality and abstractness. Quine is simply not a philosopher who would engage in substantial debates concerning, for instance, the nature of morality or the problem of free will in our deterministic world. That being said, one has good reasons to suppose that Quine considered these problems as important empirical problems, but just not problems he is particularly invested in as a professional philosopher.

Another clue to interpreting Quine on this matter is provided by the BBC television interview with Bryan Magee in 1978 (see Philosophy Overdose 2021).[11] Magee asks Quine if he includes or excludes from philosophy "the age-old questions" like "how the world got here in the first place" or "how life began". Quine quickly and resolutely answers "I exclude these from philosophy." For Quine, the question of the beginning of the world is a question for physicists and astronomists and their "conjectures" and the question how life began is a question for biologists. No surprise here. However, then Quine adds that these two questions are in fact "pseudo-questions" (after that, the camera moves to Magee and his mischievous smile). Quine perceives these questions as meaningless as he "can't imagine what an answer would look like."

According to Quine, meaningful questions and their respective answers are articulated from within a theoretical framework, so what "world"

Wittgenstein's, the task of philosophy is not to solve problems but to dissolve them by showing that there were really none there. *This doctrine has its limitations, but it aptly fits explication.* For when explication banishes a problem it does so by showing it to be in an important sense unreal; viz., in the sense of proceeding only from needless usages," (Quine 2013, 240, emphasis mine). I see this as yet another argument for interpreting Quine's view of philosophy as a very general view, adaptable to many different aims and purposes.

[11] Quine in this interview even briefly addresses the problem of free will, thanks to Magee's interest in the broader consequences of his philosophical system.

or "life" is depends on the chosen frameworks. Put another way, we cannot speak of life without already having some beliefs, i.e., assumptions, about what life is, or generally what it is to exist. In "The Limits of Knowledge," Quine writes:

> We have to work within some conceptual scheme or other; we can switch schemes, but we cannot stand apart from all of them. It is meaningless, while working within a theory, to question the reality of its objects or the truth of its laws, unless in so doing we are thinking of abandoning the theory and adopting another" (Quine 1976a, 65).[12]

Therefore, one could also say that the more internal a question to a conceptual framework is being asked, the more specific an answer will follow. Consider this example to illustrate the sense of gradability proposed here; a question "what causes genetic mutations?" will be answered from within the conceptual frameworks of genetics/evolutionary biology. For all that, the question "what is the nature of causality?" requires a more abstract answer that will not be associated with only one particular framework. One can clearly understand how the abstract answers impact the more specific ones. So, one clearly sees the place of philosophy on the continuum with science; indeed, to ask and answer abstract and general questions is a substantial contribution to (re-)wiring our web of belief, too. It could be argued that the answers to the most important general questions comprise the very wires of this web. Therefore, I claim that to deny a unique epistemic position of philosophy does not mean to deny the cognitive division of labor

[12] "Meaningless" must thus be read very loosely in this context, *and with the knowledge that it is Quine's pragmatism that displaces the residues of positivistic thinking from his philosophy.* Meaningfulness is not, for Quine, necessarily linked directly to empirically verifiable conditions (see Quine 1995b, 49). So, Quine is here simply being at one with Wittgenstein and emphasizes that some question may sound "meaningful," or even "philosophical," and yet be strictly speaking pseudo-questions in the sense that they push our language to its very limits, where it is not possible to make clear sense of our words. Let alone to specify how some of these questions are related to our technological and intellectual goals (I thank Paul Roth for forcing me to comment on this issue, and for reminding me of the parallel with Wittgenstein here).

between traditionally recognized intellectual disciplines. And we, the post-Quine philosophers, can obviously exploit this division of labor in ways that were not available or desirable to Quine.

I thus also think that Quine's understanding of naturalistic philosophy is perfectly compatible with the synoptic and synthetizing aims of philosophy as (in)famously introduced by Wilfrid Sellars: It is...the "eye on the whole" which distinguishes the philosophical enterprise" (Sellars 2007, 371). Or more forcefully and distinctively put:

> *The aim of philosophy, abstractly formulated, is to understand how things in the broadest possible sense of the term hang together in the broadest possible sense of the term.* Under "things in the broadest possible sense" I include such radically different items as not only "cabbages and kings", but numbers and duties, possibilities and finger snaps, aesthetic experience and death. To achieve success in philosophy would be, to use a contemporary turn of phrase, to "know one's way around" with respect to all these things, not in that unreflective way in which the centipede of the story knew its way around before it faced the question, "how do I walk?," but in that reflective way which means that no intellectual holds are barred (Sellars 2007, 369, emphasis mine).

For Quine, the synthetizing aspect of philosophy is inevitable. Absent a special method or the possibility to step out of all the conceptual frameworks, philosophers depend on the beliefs/knowledge generated, inter alia, by the special sciences. That is the principal assumption of Quine's intratheoretical – *from within* – approach. And, although Quine is a much more prosaic thinker than Sellars, he endorses the synoptic vision of philosophy. See again the quotes concerning philosophy above. If one reads the word "science" or the phrase "the conceptual scheme" in these passages as "the total theory of the world," one gets very close to Sellars' eye-on-the-whole designation of philosophy.

I would argue that this similarity is obscured, among many other things, by the fact that Quine often unfortunately hides in his writing style and rhetoric the assumption that he conceives science very broadly and unorthodoxly. One possible example is that, most of the time, Quine uses the

terms "natural science," "hypothetico-deductive method," "empirical checks," or "prediction" actually interchangeably. Or he also casually uses the adjectives "soft" and "hard" for sciences. But he himself does not offer any special demarcation of "natural" and "hard" science (cf., Quine 1976b, 76; Quine 1975b, 314).

In the BBC interview I referred to earlier, Magee effectively translates from Quine's austere language and takes from Quine the following synoptic message:

> So, in other words, you regard the central tasks of philosophy as the analysis and elucidation of concepts *that are central to various fields of human activity*, and also, in particular, notions like what it is for something to be a cause of something else, what it is for something to exist, what it is for something to be, shall we say, a scientific law. *The most general notions that are, as it were, the connecting tissue of thought and that we have to use and have to employ in the specific activities that people like scientists – or it could even be politicians, lawyers and so on – are engaged in.* Is that a correct way of putting your view?
>
> [Quine laconically replies:] Yes, yes. I would agree with that. (Philosophy Overdose 2021, transcribed and emphasized by me)

In light of what has been said so far, I claim that even Quine is the child of the 20th century (philosophy of) science when it comes to the vocabulary.

2.2 Science is Not the Only Game in Town

If one also seriously takes into account Quine's inclination to view science as a Wittgensteinian language game – one among several such games – it follows that other human linguistic practices, such as poetry and fiction writing, exemplify other and legitimate language games. Additionally, for Quine, our modern science game has already moved past its *primary* evolutionary purpose of predicting our experiences, and its two dominant purposes are currently "technology" and "understanding," meaning specific pragmatic purposes and specific intellectual purposes (Quine 1992, 20). I will put aside the technological purposes, as I think no one in their right mind would deny the technological advances of our present-day sciences.

However, if one takes on board the plurality of language games, one is entitled to ask how to account, in the Quinean spirit, for the difference between "understanding" from within science and, for instance, from within fiction writing (and, for that matter, from within any kinds of writings that were deemed non-scientific at the time). If fiction qua a form of art also "teaches" and "enlightens" us, as Quine unproblematically holds (Quine & Ullian 1978, 4), when should we adhere to science and when, for instance, to a novel? Again, these sorts of questions are off Quine's radar most of the time, so I am aware that my answers cannot be supported by Quine's explicit reflections.

However, I am not particularly interested in interpreting Quine as in interpreting his naturalistic legacy. Put another way, from now on, I engage in the Quinean naturalistic explication, without necessarily arguing that it is an explication Quine himself would have defended. The main purpose of this explication is to clarify the relationship between science, the arts, and various forms of second-order inquiry – often referred to as "philosophy" but also by other names, such as "ethics." Fiction writing thus represents the limit case, as Quine does not go so far as to claim that even fiction is *cognitively* worthless. So, secondarily, the explication should clarify why and in what sense Quine does *not* have scientistic inclinations.

I will start in a roundabout way by deliberating on the difference between factual and non-factual discourse. Then, I will briefly discuss fictional discourse in the artistic sense. Stephen Jay Gould argued some time ago that, if perceived properly, science and religion are not in any conceptual or explanatory conflict.

> No such conflict should exist because each subject has a legitimate magisterium, or domain of teaching authority – and these magisteria do not overlap (the principle that I would like to designate as NOMA, or "nonoverlapping magisteria"). *The net of science covers the empirical universe: what is it made of (fact) and why does it work this way (theory). The net of religion extends over questions of moral meaning and value. These two magisteria do not overlap, nor do they encompass all inquiry (consider, for starters, the magisterium of art and the meaning of beauty* (Gould 2011a [1997], 274, emphasis mine).

> The attainment of wisdom in a full life requires extensive attention to both domains – *for a great book tells us that the truth can make us free and that we will live in optimal harmony with our fellows when we learn to do justly, love mercy, and walk humbly* (ibid., 271, emphasis mine).

As an evolutionary scientist and left-wing liberal, Gould was sympathetic to many kinds of intellectual and ideological movements; Gould understood very well that to agree with NOMA is easier than to actually practice it. The vast majority of questions central to our everyday lives cannot be moved too far away from the boundaries of the given magisteria. In this respect, for Gould, the friction surfaces between science and religion represent the paradigm of NOMA in practice. Therefore, although Gould primarily speaks of science and religion, he undoubtedly intends to extrapolate from these two to say something general about our human epistemic condition. In this sense, I regard NOMA as a naturalistic thesis.[13]

> *Many of our deepest questions call upon aspects of both for different parts of a full answer – and the sorting of legitimate domains can become quite complex and difficult.* To cite just two broad questions involving both evolutionary facts and moral arguments: Since evolution made us the only earthly creatures with advanced consciousness, what responsibilities are so entailed for our relations with other species? *What do our genealogical ties with other organisms imply about the meaning of human life?* (ibid., 274, emphasis mine).

Now, at face value, Quine and all like-minded naturalists would certainly agree that science is a theory-driven, fact-stating discourse (a language game). Similarly, naturalists have no problem recognizing what Gould hints at here; scientific discourse does not, in itself, imply any normative guidelines. By explaining our origins in evolutionary terms, scientists are not prescribing anything. Their explanations do not directly translate into

[13] I use "naturalistic" here in its most basic sense, referring simply to the aim of reconfiguring the relationship between scientific discourse and other forms of discourse. However, I consider this sense to be a properly understood interpretation of Quine's perspective as well (see again, e.g., Quine 1995a).

value-stating discourse; in other words, these explanations are not direct value claims. The "ought" cannot be derived from the "is" directly without further background assumptions and aims, such as chosen ethical frameworks or societal values that guide and ground our interpretations.

For the type of holistic naturalism I derive from Quine, this follows, inter alia, since the very constituting and communicating "facts" are already inseparable from our historical cultures/scientific communities, i.e., inseparable from many other (background) beliefs, interests, and purposes. Let alone any further attempts to explain or theorize the facts. "Only theologians and unthinking ideologues imagine that normative conclusions can be read off some recitation of the facts" (Roth in Domańska et al. 2019, 544; cf., Gould 2011a [1997], 282).

At any rate, Quine and Gould are in agreement that "science" has a monopoly in determining what counts as facts, i.e., monopoly in generating meaningful fact-stating statements. This monopoly, however, is not tied to any methodological ownership. Nor to any unity-of-science trusts. In Quine's view, science has no ahistorical demarcation criterion from other commonly recognized *non-fictional* practices and discourses. The monopoly is rather grounded in the appreciation that the general intersubjective criteria (of individual sciences) has been the most successful tool for settling, i.e., rationalizing, our beliefs about the world, and the most effective tool for manipulating the world.

So, the claim so far is that as long as religion/theology, or any other non-fictional discourse for that matter, does not claim to be more fundamental or accurate than science with respect to factual matters, no conflict arises. For the naturalist of the Quinean kind, this does not necessarily mean that theology must abruptly accept, e.g., all the conclusions and conjectures of up-to-date physics. In this regard, respect for the facts is enough, regardless of how loosely one demarcates them. Take as an example the concept of "soul." This implies that understanding the concept of the soul dualistically (as a sui generis non-material entity) would need to be based on some form of intersubjective criterion and/or epistemologically-pragmatic purposes. The dualism of body and soul could, in principle, be empirically vindicated if it would give us some explanatory or technological benefits (for example, in sociology, medicine, etc.). The

same would hold for concepts such as "prayer," "miracle," "revelation," or even "God."

Gould, however, seems to primarily target an additional issue; the nature of non-factual discourse, i.e., typically ethical/value discourse. As he further clarifies:

> NOMA also cuts both ways. If religion can no longer dictate the nature of factual conclusions properly under the magisterium of science, *then scientists cannot claim higher insight into moral truth from any superior knowledge of the world's empirical constitution. This mutual humility has important practical consequences in a world of such diverse passions* (Gould 2011a [1997], 281–282, emphasis mine).

It is obvious that theology exemplifies only one of the possible ethical and value discourses in Gould's understanding of NOMA (cf., Gould 2011b [1999]). So, if anything, Gould can be charged with the intention of narrowing down the role of religion and theology exclusively to their ethical dimensions, and interpreting the existential (ipso facto factual) statements of theology in a *non-literal* way. But that is a problem for another day. Now, against this background, instead of science and religion, one could just speak of fact-stating discourse and ethical/value discourse, or generally of factual and non-factual discourse.

NOMA then certainly does not imply, as for example Richard Dawkins (1998) incorrectly infers, that the non-factual (ethical) discourse of theology can "unproblematically" work with *transcendental* concepts or notions (e.g., heaven, life after death, etc.). That would indeed amount to making *super*-natural existential claims, however implicitly. On the other hand, NOMA does not prevent us from working with transcendental concepts in metaphorical ways, e.g., operationalizing transcendental concepts qua Kantian regulative ideas.

This fact also seems to be one of the main motivations for Gould when introducing NOMA:

> While I cannot personally accept the Catholic view of souls, *I surely honor the metaphorical value of such a concept both for grounding moral discussion and for expressing what we most value about human potentiality: our decency, care, and all the ethical and intellectual struggles that the evolution of consciousness*

> *imposed upon us* (Gould 2011a [1997], 282, emphasis mine, cf., 271).

Interestingly, Quine, when parsing the notion of the "absolute truth," also states that "metaphor is perhaps a handy category in which to accommodate transcendental concepts, from a naturalistic point of view" (Quine 1995a, 261). In general, it could be said that *non*-factual use of language simply serves different purposes and tackles different problems in which one is interested. It addresses different kinds of questions, e.g., "what is worth living for?," or "why should I stay healthy?" Therefore, even though Rorty actually thinks otherwise, I argue that Quine allows for considering our diverse uses of factual as well as non-factual language as tools, or in Rorty's own words, "techniques of problem solving" (Rorty 2021, 184). On my reading then, Quine's *functional* view of our human discoursing also leads to the pragmatist stance Rorty himself sharply depicts this way:

> Democritus, Newton, and Dalton solved some problems with particles and laws. Darwin, Gibbon, and Hegel solved others with narratives. Carpenters solve others with hammers and nails, and soldiers still others with guns. *Philosophers' problems are about how to prevent the words used by some of these problem-solvers from getting in the way of other words used by other problem-solvers* (ibid., emphasis mine; cf., Thomasson 2024 for a more focused take on this matter).

I am also convinced that the functional view of discourse gives us a suitable conceptual bridge for comprehending the cognitive component of the arts. Due to space constraints, I briefly focus on fiction writing/novels. It is common sense that fiction writers can always use the language to confront us with "what ifs" or "as ifs" propositions related to our "real" lives. They can always imaginatively (re-)construct dialogues between two "real" persons or simply compose entirely fictional worlds and events to educate and/or entertain us. Also, this nicely illustrates that the first-person perspectives, although strictly speaking not scientifically useful in the naturalistic sense, can often be valid and valuable with respect to our human needs, interests, and purposes.

So far so good, but one might still wonder how to satisfactorily differentiate between artistic, and in some sense scientific usage of language.

Take, for example, someone like Kendall L. Walton and his methodically naturalistic explicatory attitude towards all the art forms. Walton conceives of the artistic modes of expression as "games of make-believe:" "I take seriously the association with children's games – with playing house and school, cops and robbers, cowboys and Indians, with fantasies built around dolls, teddy bears, and toy trucks. We can learn a lot about novels, paintings, theater, and film by pursuing analogies with make-believe activities like these" (Walton 1990, 4).

Construed as such, the games of make-believe cannot be limited only to arts (or solely to discursive activities), but are rather

> *a pervasive element of human experience*.…There is nothing distinctively "aesthetic" about make-believe itself at all.…I suspect that make-believe may be crucially involved as well in certain religious practices, in the role of sports in our culture, in the institution of morality, *in the postulation of "theoretical entities" in science, and in other areas in which issues of metaphysical "realism" are prominent*… (ibid., 7, emphasis mine).

This approach is moreover congenial to Quine's metaphysically *anti-representational* view of reality. As I have already established, Quine often stresses that the objects science postulates are, from the epistemological vantage point, useful myths or fictions (see again the previous section). And it also seems to be a quite expectable conclusion when the demarcation criterion of science is being abandoned.

Therefore, even though one cannot hope to identify a universal and unambiguous demarcation between scientific works and, e.g., novels, this does not mean that there are not pragmatically useful criteria for adjudicating individual instances. As naturalism dictates; the demarcation of scientific propositions will always have to be done empirically and intratheoretically, i.e., always ad hoc. Walton himself offers an illuminating exemplification of this type of constraint. Consider in this light the following passages:

> An important symptom of the difference between *The Origin of Species* and works like *Gulliver's Travels*…is that what is said in *The Origin of Species* does not of itself warrant assertions like "Species evolved by means of natural selection." *It justifies such assertions only insofar as it provides good reason to think they*

> *are true*. But the sentences in *Gulliver's Travels* warrant the assertive utterance "A war was fought over how to break eggs," *quite apart from whether they give us reason to think such a war actually was fought* (Walton 1990, 71, emphasis mine).
>
> If we are to believe the theory of evolution, it is because that theory is true, or because there is good evidence for it, not because it is expressed in *The Origin of Species* – although of course *The Origin of Species* might convince us of the theory's truth or inform us of evidence for it. *Darwin's book itself does not prescribe believings. So we cannot conclude that it prescribes imaginings, even if believing involves imagining* (ibid., 70–71, emphasis mine).

Consequently, the very purposes of scientific works and novels can be, in principle, very similar or even identical. They can aim at the same goals via different means. Again, no principled reasons exist for identifying the universal purposes of either scientific or artistic works.[14] Whether or not a statement has a scientific status must be decided against the background of other beliefs; it must be done from within our web of belief, to use the Quinean expression once again. Let me return to the example of the novel; it is obvious that every writer works, explicitly or implicitly, with the purposes of their work in manifold, and in principle, open-ended ways. And

[14] Especially not for evaluating single, isolated statements. It is also a dialectical process; dispositions as well as expectations of the audience matter. Walton addresses this point via speech-act theory: "*The fundamental disanalogy between illocutionary actions and acts of fiction making comes out in differences in the roles of agents' intentions*. A crucial question for a person on the receiving end of an illocutionary action is almost always, Did he mean it? Did he intend to assert this, to promise that, to issue such and such an order or apology? But one may well read a story or contemplate a (fictional) picture without wondering which fictional truths the author or artist meant to generate. Photographers, especially, can easily be unaware of fictional truths generated by their works. Authors and other artists may be surprised at where extrapolation from the fictional truths they intentionally generated leads. *This need not make any particular difference to the appreciator – unless he is concerned with what the artist might be asserting in producing the fiction, what illocutionary actions she might be performing in the process of, and in addition to, producing it*" (ibid., 88, emphasis mine).

these ways can intersect (or be wholly interwoven) with other purposes of other intellectual endeavors and disciplines. It seems to be an understandable consequence of our human condition and of our linguistic modalities. To give one specific example, Milan Kundera in his *The Art of the Novel* explicates the "spirit of the novel" at one point as follows:

> *As a model of [the] Western world, grounded in the relativity and ambiguity of things human, the novel is incompatible with the totalitarian universe.* This incompatibility is deeper than the one that separates a dissident from an apparatchik, or a human-rights campaigner from a torturer, because it is not only political or moral but ontological. By which I mean: *The world of one single Truth and the relative, ambiguous world of the novel are molded of entirely different substances. Totalitarian Truth excludes relativity, doubt, questioning; it can never accommodate what I would call the spirit of the novel* (Kundera 2020, 10, long emphases mine).

Kundera also thinks that the novel is inherently historical. "Novels take place in a time that has a date and is thoroughly historical," (ibid., 20). But one must carefully distinguish between "*the historical dimension of human existence*," and "*the illustration of a historical situation*;" meaning "popularizations that translate non-novelistic knowledge into the language of the novel." Kundera then concludes: "Well, I'll never tire of repeating: The novel's sole *raison d'etre* is to say what only the novel can say" (ibid.). So, for him, there are several principles distinguishing novels from scientific (e.g., historiographic) works. One of them being a different point of focus:

> *Historiography writes the history of society, not of man. That is why the historical events my novels talk about are often forgotten by historiography.* Example: In the years that followed the 1968 Russian invasion of Czechoslovakia, the reign of terror against the public was preceded by officially organized massacres of dogs. *An episode totally forgotten and without importance for a historian, for a political scientist, but of the utmost anthropological significance!* By this one episode alone I suggested the historical climate of *The Farewell Party.* Another example: At the crucial point of Life Is Elsewhere, History intervenes in the form of an

> inelegant and shabby pair of undershorts; there were no others to be had at the time; faced with the loveliest erotic occasion of his life, Jaromil, for fear of looking ridiculous in his shorts, dares not undress and takes flight instead. *Inelegance! Another historical circumstance forgotten, and yet how important for the person obliged to live under a Communist regime* (ibid., 21, long emphases mine).

This is one of the many possible approaches, compatible with the Quinean approach I develop and defend here, to distinguishing between the purposes and cognitive components of scientific works (in our case above historiographical works), and novels (in our case above of novels dealing with the history of humanity).

When taken together, it represents an inquiry into the human "soul" in a manner that would be appreciated by not only Gould but Quine as well. True, Quine writes at the very end of his autobiography: "*I have little bent for soul-searching.*...My way of coping with the spells of nostalgia, loneliness, anxiety, or boredom over the years has been to escape into my projects" (Quine 1985, 475, emphasis mine).[15] However, his naturalizing project, which emphasizes the role of the third-person perspective over the first-person one, nevertheless creates a space that preserves all the typically acknowledged and appreciated purposes and cognitive functions of the arts, or so I have argued with the assistance of Walton and Kundera above.

To conclude, in this section, I have proposed that Quine's functional view of our language use (hence also scientific, philosophical, ethical, and artistic discourse) empowers naturalists to see human linguistic practices as tools for specific pragmatic and intellectual purposes. Such a view also

[15] Quine then continues: "I am deeply moved by occasional passages of poetry, and so, characteristically, I read a little of it. I respond similarly to passages of grand opera, and this is due to the libretto as much as to the music. Otherwise I have a poor memory for fiction, for it resists integration with my system of the world. I appreciate style more than plot. I enjoy the eighteenth-century brand of humor that we get in Fielding's prefaces and belatedly in Dickens; also the humor of W. S. Gilbert and Wodehouse and, in moderation, zany S. J. Perelman. Encouraged by the indulgent interest of readers who have read all the way to here, I am evidently accomplishing some soul-searching after all" (ibid., 476).

sheds light on the relationship between and different functions of third-person perspective (paradigmatic for but not exclusive to science); and first-person perspective (utilized often but not uniquely within the arts).

In addition, my own attempt to interpret Quine on this matter exemplifies a case of philosophical exemplification under natural*ism*.

3. Resumé

The first section of this essay gave an explication of Quine's hint that, under naturalism, realism and idealism are dissolved into each other. The second section explicated Quine's dictum that (naturalistic) philosophy is continuous with (natural) science. As I reveal, both these motifs are dealt with by Quine's characteristic *from within* approach, i.e., they are dealt with by emphasizing the intra-theoretic and intra-linguistic nature of all human empirical inquiry. As Quine himself always accepted: "I philosophize from the vantage point only of our own provincial conceptual scheme and scientific epoch, true; but I know no better" (Quine 1969d: 25). Such an approach is the hallmark of the Quinean type of naturalism, and so this essay attempted to further clarify it in concrete and contextual terms.

Additionally, since Quine is commonly considered to be an intellectual father of all current forms of naturalism, my partially historical and partially conceptual exegesis of his views in this essay can also further contribute to a better understanding of naturalism and pragmatism of today.

Acknowledgements

I am indebted to Paul A. Roth and Willem A. deVries for their insightful comments on earlier versions of this essay. I also wish to thank the anonymous reviewer, whose careful reading and valuable feedback significantly improved the penultimate version.

Funding

My work on this essay was made possible thanks to the internal grant "Truth Wins, or Winner's Truth? The Problem of Truth in Contemporary Philosophy" (2022–2024), awarded by the University of Hradec Králové, Philosophical Faculty.

References

Borradori, Giovanna. 2008. *The American Philosopher: Conversations with Quine, Davidson, Putnam, Nozick, Danto, Rorty, Cavell, MacIntyre, Kuhn*, translated by Crocitto, R. Chicago: University of Chicag Press.

Bryant, Amanda. 2020. "Naturalisms." *Think*, 19(56): 35–50. https://doi.org/10.1017/S1477175620000196

Danto, Arthur. C. 1967. "Naturalism." In *The Encyclopedia of Philosophy*, volume 5, edited by Edwards, 448–50. New York: The Macmillan Company and The Free Press.

Davidson, Donald. 1973–1974. "On the Very Idea of a Conceptual Scheme." *Proceedings and Addresses of the American Philosophical Association*, 47: 5–20. https://doi.org/10.2307/3129898

Davidson, Donald. 2002 (1970). "Mental Events." In *Contemporary Materialism: A Reader*, edited by P. K. Moser and J. D. Trout, 122–37. London: Routledge.

Dawkins, Richard. 1998. "When Religion Steps on Science's Turf: The Alleged Separation between the Two is Not So Tidy." *Free Inquiry*, (18)2: 18–19.

De Caro, Mario. and David Macarthur (eds.). 2004. "Naturalism in Question." Cambridge, Massachusetts: Harvard University Press. https://doi.org/10.2307/j.ctv22jnrf5

Dewey, John. 1911. "Brief Studies in Realism. II." *The Journal of Philosophy, Psychology and Scientific Methods*, 8(20): 546–54.

Dreben, Burton. 1992. "Putnam, Quine – and the Facts." *Philosophical Topics*, 20(1): 293–315. https://doi.org/10.5840/philtopics199220110

Gould, Stephen Jay. 2011a. (1997). "Non-Overlapping Magisteria." In *Leonardo's Mountain of Clams and the Diet of Worms: Essays on Natural History*, 269–84. Cambridge, Massachusetts: Harvard University Press.

Gould, Stephen Jay. 2011b. (1999) *Rocks of Ages: Science and Religion in the Fullness of Life*. New York: Ballantine Books.

Heath, Joseph. and Catherine Rioux. 2018. "Recent Trends in Evolutionary Ethics: Greenbeards!" *Biology & Philosophy*, 33(1): 1–20. https://doi.org/10.1007/s10539-018-9627-1

Hofstadter, Albert. 1954. "The Myth of the Whole: A Consideration of Quine's View of Knowledge." *The Journal of Philosophy*, 51(14): 397–417.

Honderich, Ted. (ed.). 1995. *The Oxford Companion to Philosophy*. Oxford: Oxford University Press.

Jaksland, Rasmus. 2024. "Naturalized Metaphysics in the Image of Roy Wood Sellars and not Willard Van Orman Quine." *Metaphilosophy*, 55(2): 214–30. https://doi.org/10.1111/meta.12677

Keskinen, Antti. 2012. "Quine on Objects: Realism or Anti-Realism?" *Theoria*, 78(2): 128–45. https://doi.org/10.1111/j.1755-2567.2012.01141.x

Kundera, Milan. 2020. *The Art of the Novel*. London: Faber & Faber [Ebook].

Menand, Louis. 2002. *The Metaphysical Club: A Story of Ideas in America*. New York: Macmillan.

Mitrović, Branko. 2022. "A Naïve Realist Rumination on the Roth-and-Dewulf Versus Currie-and-Swaim Exchange." *Journal of the Philosophy of History*, 16(3): 295–307. https://doi.org/10.1163/18722636-12341482

Papineau, David. 2021. "Naturalism," *The Stanford Encyclopedia of Philosophy*. Edited by E. N. Zalta. https://plato.stanford.edu/archives/spr2021/entries/naturalism/

Peregrin, Jaroslav. 1999. *Zen & The Art of (Davidsonian) Analytic Philosophy. Jarda*. http://www.jarda.peregrin.cz/mybibl/PDFTxt/476.pdf

Peregrin, Jaroslav. 2023. *Normative Species: How Naturalized Inferentialism Explains Us*. New York: Routledge.

Philosophy Overdose. 2021, December 16. "The Ideas of Quine – Bryan Magee & Quine" (1978) [Video]. YouTube. https://www.youtube.com/watch?v=NnhUzp1txtk&t=2s

Pils, Raimund. 2020. "Quine's Scientific Realism Revisited." *Theoria*, 86(5): 612–42. https://doi.org/10.1111/theo.12273

Price, Huw. 2011. *Naturalism Without Mirrors*. New York: Oxford University Press.

Quine, Willard Van Orman. 1957. "The Scope and Language of Science." *The British Journal for the Philosophy of Science*, 8(29): 1–17.

Quine, Willard Van Orman. 1961a. (1951). "Two Dogmas of Empiricism." In *From a Logical Point of View*: *Logico-Philosophical Essays*, 20–46. New York: Harper and Row, Publishers.

Quine, Willard Van Orman. 1961b. (1948) "On What There Is." In *From a Logical Point of View*: *Logico-Philosophical Essays*, 1–19. New York: Harper and Row, Publishers.

Quine, Willard Van Orman. 1961c. (1950). "Identity, Ostension, and Hypostasis." In *From a Logical Point of View: Logico-Philosophical Essays*, 65–79. New York: Harper and Row, Publishers.

Quine, Willard Van Orman. 1969a. "Natural Kinds." In *Ontological Relativity*, 114–38. New York: Columbia University Press.

Quine, Willard Van Orman. 1969b. "Epistemology Naturalized." In *Ontological Relativity*, 69–90. New York: Columbia University Press.

Quine, Willard Van Orman. 1969c. "Ontological Relativity." In *Ontological Relativity*, 26–68. New York: Columbia University Press.

Quine, Willard Van Orman. 1969d. "Speaking of Objects." In *Ontological Relativity*, 1–25. New York: Columbia University Press.

Quine, Willard Van Orman. 1975a. (1969). "Replies" In *Words and Objections*, edited by D. Davidson and J. Hintikka, 292–352. Dotrecht: D. Reidel.

Quine, Willard Van Orman. 1975b. "On Empirically Equivalent Systems of the World." *Erkenntnis*, 9(3): 313– 28.

Quine, Willard Van Orman. 1976a. "The Limits of Knowledge." In *The Ways of Paradox and Other Essays*, 59–67. Revised and Enlarged Edition. Cambridge, Massachusetts: Harvard University Press.

Quine, Willard Van Orman. 1976b. "Necessary Truth." In *The Ways of Paradox and Other Essays*, 68–76. Revised and Enlarged Edition. Cambridge, Massachusetts: Harvard University Press.

Quine, Willard Van Orman. 1981a. "Things and Their Place in Theories." In *Theories and Things*, 1–23. Cambridge, Massachusetts: Harvard University Press.

Quine, Willard Van Orman. 1981b. "Fives Milestones of Empiricism." In *Theories and Things*, 67–72. Cambridge, Massachusetts: Harvard University Press.

Quine, Willard Van Orman. 1984. "Relativism and Absolutism." *The Monist*, 67(3): 293–96.

Quine, Willard Van Orman. 1985. *The Time of My Life: An Autobiography*. Cambridge, Massachusetts: MIT Press.

Quine, Willard Van Orman. 1987. *Quiddities*: *An Intermittently Philosophical Dictionary*. Cambridge, Massachusetts: Harvard University Press.

Quine, Willard Van Orman. 1992. (1990). *Pursuit of Truth*. 2nd edition. Cambridge, Massachusetts: Harvard University Press.

Quine, Willard Van Orman. 1995a. "Naturalism; Or, Living within One's Means." *Dialectica*, 49(2–4): 251–63.

Quine, Willard Van Orman. 1995b. *From Stimulus to Science*. Cambridge, Massachusetts: Harvard University Press.

Quine, Willard Van Orman. 1996. "Progress on Two Fronts." *The Journal of Philosophy*, 93(4): 159–63.

Quine, Willard Van Orman. 2008. "The Innate Foundational Endowments." In *Confessions of a Confirmed Extensionalist and Other Essays*, edited by D. Føllesdal and D. B. Quine, 176–81. Cambridge, Massachusetts: Harvard University Press.

Quine, Willard Van Orman. 2013. (1960). *Word and Object*. Cambridge, Massachusetts: MIT press. https://doi.org/10.7551/mitpress/9636.001.0001

Quine, Willard Van Orman and Joseph. S. Ullian. 1978. (1970). *The Web of Belief*. 2nd edition. New York: Random

Rorty, Richard. 1976. "Realism and Reference." *The Monist*, 59(3): 321–40.

Rorty, Richard. 1998. *Truth and Progress: Philosophical Papers*, volume 3. New York: Cambridge University Press.

Rorty, Richard. 2021. *Pragmatism as Anti-Authoritarianism*. Cambridge, Massachusetts: Harvard University Press. https://doi.org/10.4159/9780674270077

Rosenberg, Alex. 1996. "A Field Guide to Recent Species of Naturalism." *The British Journal for the Philosophy of Science*, 47(1): 1–29. https://doi.org/10.1093/bjps/47.1.1

Rosenberg, J. F. 2007. *Wilfrid Sellars: Fusing the Images*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780199214556.003.0002

Roth, Paul A. 2003. "Why There Is Nothing Rather Than Something: Quine on Behaviorism, Meaning, and Indeterminacy." In *Philosophy, Psychology, and Psychologism: Critical and Historical Readings on the Psychological Turn in Philosophy*, edited by D. Jacquette, 263–87. Dordrecht: Springer. https://doi.org/10.1007/0-306-48134-0_13

Roth, Paul A. 2020. *The Philosophical Structure of Historical Explanation*. Evanston, Illinois: Northwestern University Press.

Roth, Paul A. 2023. "What Does Translation Translate? Quine, Carnap, and the Emergence of Indeterminacy." In *The Philosophical Project of Carnap and Quine*, edited by S. Morris, 154–76. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108664202.012

Sellars, Wilfrid. 2007. (1962). "Philosophy and the Scientific Image of Man." In *In the Space of Reasons: Selected Essays of Wilfrid Sellars*, edited by R. Brandom and K. Scharp, 369–408. Cambridge, Massachusetts: Harvard University Press.

Thomasson, Amie. L. 2024. "How Should We Think about Linguistic Function?" *Inquiry*, 67(3): 840–71. https://doi.org/10.1080/0020174X.2022.2074886

Tuomisto, Hanna, Matleena Tuomisto, and Jouni T. Tuomisto. 2018. "How Scientists Perceive the Evolutionary Origin of Human Traits: Results of a Survey Study." *Ecology and Evolution*, 8(6): 3518–33. https://doi.org/10.1002/ece3.3887

Walton, Kendall. L. 1990. *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Cambridge, Massachusetts: Harvard University Press.

Wright, Crispin. 1993. *Realism, Meaning, and Truth*. 2nd ed. Cambridge, Massachusetts: Blackwell. https://doi.org/10.2307/2219715